

**INFORME TÉCNICO – ENTREGA FINAL****CONCEPTUALIZACIÓN DEL PROBLEMA**

En un entorno empresarial colombiano marcado por la hipercompetencia, la saturación de oferta y la transformación digital, la capacidad de una organización para entender y anticiparse a las necesidades del cliente se ha vuelto una necesidad estratégica. El problema central que se aborda en este reto es la búsqueda de una solución automatizada para suplir la limitada capacidad de las empresas para personalizar la experiencia del consumidor. Esta problemática compromete no solo la satisfacción y fidelización del cliente, sino también el aprovechamiento pleno del valor de los datos transaccionales acumulados, generando ineficiencias operativas y oportunidades de venta desaprovechadas. En este contexto, la ausencia de un enfoque inteligente en la recomendación de productos impide la consolidación de relaciones duraderas con los usuarios y dificulta la optimización de portafolios en función de patrones reales de comportamiento y alineación estratégica.

La problemática tratada también trasciende el ámbito tecnológico para situarse en una intersección entre la analítica de datos, estrategia comercial y diseño de experiencias. Por ende, se plantea la necesidad de desarrollar un sistema robusto que articule capacidades de *machine learning* con objetivos de negocio claramente definidos: incrementar la conversión de ventas, facilitar la labor de los asesores en punto de venta y generar *insights* accionables para la toma de decisiones. La solución debe ser capaz de absorber, interpretar y capitalizar datos heterogéneos de distintas fuentes (transacciones, cotizaciones y ventas B2B), para derivar recomendaciones personalizadas que no solo respondan al historial del cliente, sino que estén alineadas con el portafolio estratégico de la compañía. En suma, se trata de superar las limitaciones actuales de la gestión comercial a través de un modelo de recomendación inteligente que funcione como catalizador de valor integral para el negocio.

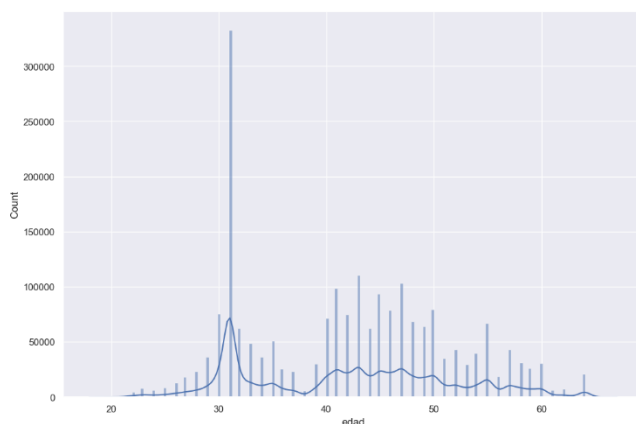
**ANÁLISIS DESCRIPTIVO**Base Transaccional B2C (base 1 transaccional)

Para comenzar el desarrollo del proyecto, se llevó a cabo un análisis exploratorio detallado utilizando la base de datos B2C, compuesta por 2.099.836 registros y 18 variables explicativas. Esta base contiene información clave sobre el comportamiento de compra del consumidor final, incluyendo datos como el tipo de producto, su categoría y subcategoría, la fecha de transacción, el valor total de la compra y su alineación con el portafolio estratégico B2C.

La calidad de los datos fue satisfactoria en general. No se detectaron registros duplicados y el porcentaje de datos faltantes no fue superior al 2% en la variable precio y al 0.008% en la

variable zona, estos problemas de completitud fueron resueltos eliminando dichos registros ya que la ausencia de esas dos variables los convierte en información no relevante para el proyecto. Con respecto a la preparación de los datos, la variable “fecha\_factura” fue transformada correctamente al formato *datetime*, lo que facilita la exploración de tendencias temporales y la identificación de patrones estacionales.

En cuanto a las variables numéricas, la variable “valor” presentó una distribución claramente sesgada hacia la derecha. El 75% de las transacciones registran valores iguales o inferiores a aproximadamente 37 unidades monetarias, lo que indica una alta concentración de compras de bajo valor. Sin embargo, se observan valores extremos que alcanzan hasta 56,876 unidades, lo cual sugiere la existencia de adquisiciones puntuales de alto valor, posiblemente relacionadas con productos premium, compras por volumen o campañas especiales. La mediana se sitúa en 13.32 unidades, reflejando una diferencia significativa frente al valor máximo y reafirmando el sesgo positivo. Por otro lado, la variable “edad” muestra una distribución relativamente simétrica, con una media de aproximadamente 41.9 años y una mediana cercana (43 años). Sin embargo, hay una altísima presencia de compradores al inicio de sus 30, estos podrían representar a los clientes que están independizando y por ende realizan muchas compras para la adecuación de su nuevo hogar.



*Ilustración 1 Histogramas de frecuencia de edad*

La variable “alineación con portafolio estratégico” mostró una alta dispersión, con valores entre -24,187 y 4,162. Aunque la media (3.95) y la mediana (1.41) indican una leve alineación positiva en la mayoría de las transacciones, la presencia de valores negativos y extremos sugiere comportamientos comerciales dispares. Esto resalta la necesidad de guiar a los clientes hacia productos más alineados con la estrategia de la compañía.

El análisis de correlación reveló una relación moderada y positiva entre el “valor” y la “alineación con portafolio estratégico” ( $r = 0.53$ ), lo que sugiere que las transacciones de mayor valor tienden a estar más alineadas con los objetivos estratégicos de la compañía. También se observa una correlación moderada entre “valor” y “cantidad” ( $r = 0.51$ ), indicando que el monto total está fuertemente influenciado por el volumen de productos adquiridos.

En general, el resto de las variables presentan correlaciones débiles o casi nulas entre sí, lo que indica una baja multicolinealidad y una relativa independencia entre los atributos.

En lo que respecta a las variables categóricas, se evidencia una fuerte concentración de transacciones en un número reducido de municipios y zonas. El municipio de Curití representa más del 27% del total de transacciones, seguido por Natagaima y Villanueva. A nivel de zona, Santander domina ampliamente con más de 700 mil registros, seguido de Tolima y Antioquia. Esta concentración geográfica plantea oportunidades para diversificar la cobertura comercial y fortalecer estrategias locales en zonas con menor participación.

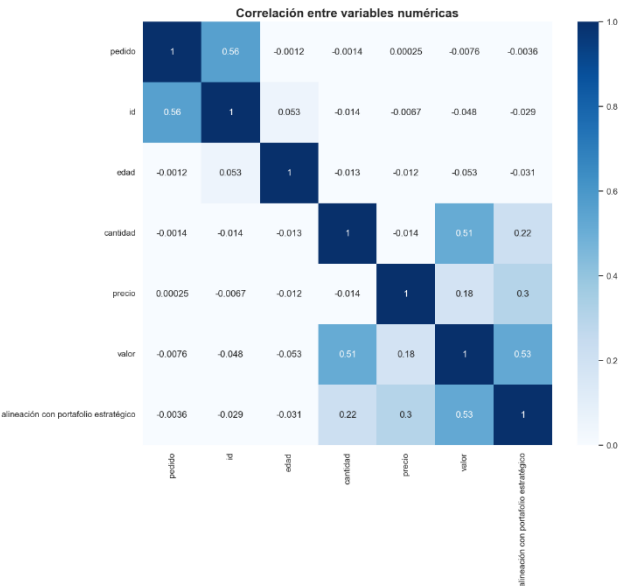


Ilustración 2 Correlación de variables numérica

En cuanto a los puntos de venta y asesores, los cinco principales puntos concentran una proporción significativa del volumen total, con “punto\_venta\_7” a la cabeza (132,386 transacciones).

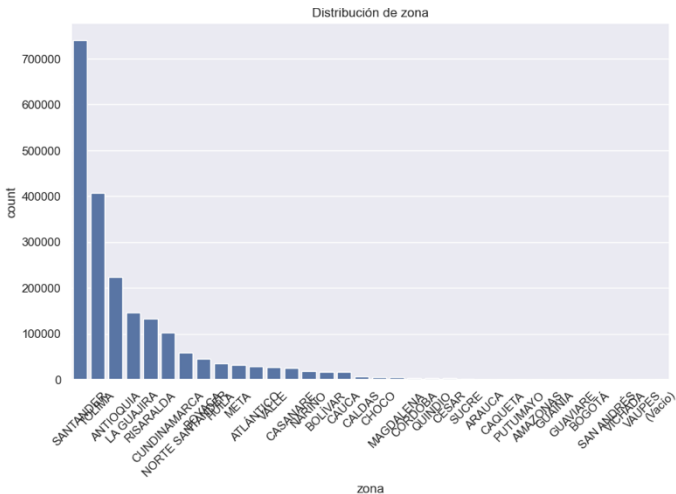


Ilustración 3 Frecuencia de departamentos de los registros

mercado dominada por unos pocos perfiles de tienda. A nivel de producto, se identificaron más de 7,200 referencias únicas, aunque el producto más vendido (“producto\_49”) representa apenas el 2.56% del total, lo que indica una alta dispersión en las preferencias del consumidor. Este patrón se mantiene a nivel de subcategorías y categorías, con más de 100 subcategorías y 27 categorías, pero una fuerte concentración en las más populares como “subcategoría\_5” y “categoría\_3”.

De manera similar, algunos asesores como “asesor\_137” y “asesor\_7” superan las 14 mil ventas, mientras que más de 300 asesores tienen apenas una transacción registrada, lo que refleja una distribución altamente desigual en el desempeño comercial.

Respecto a los clústeres, más del 80% de las ventas están concentradas en “cluster\_tienda\_3” y “cluster\_tienda\_2”, lo cual indica una segmentación de

En cuanto a las categorías macro, cinco agrupan la totalidad de las transacciones, destacándose “categoría\_macro\_2” con más de 1.3 millones de registros (62% del total), seguida por “categoría\_macro\_4” y “categoría\_macro\_1”. Finalmente, la variable “color” también muestra un patrón concentrado: más del 47% de las referencias no tienen color identificado, y entre las restantes predominan tonos neutros como gris, blanco y beige. Estos

hallazgos destacan la necesidad de mejorar la calidad del dato en ciertas dimensiones y de adoptar estrategias de diversificación que reduzcan la dependencia de territorios, asesores y productos específicos.

#### Base de Cotizaciones B2C (base 2 cotizaciones)

Para comprender la dinámica de cotización del segmento B2C, se llevó a cabo un análisis exploratorio exhaustivo sobre la base de datos proporcionada, la cual contiene 180.387 registros distribuidos en 11 variables explicativas. Esta base refleja el comportamiento de los clientes en procesos de solicitud de información o intención de compra, permitiendo identificar patrones de interés sobre el portafolio comercial y la interacción con los productos ofrecidos. Con respecto a la calidad de datos se encontró que la base de datos estaba completa, sin embargo, se evidenció un problema de duplicados que fue resuelto eliminándolos. La consistencia de las variables numéricas y categóricas, junto con la validez de estos fueron coherentes con el contexto y los datos. Los datos de fecha no estaba en formato *datetime* lo que podía perjudicar el manejo de los mismo y por ende se cambió al formato esperado.

El análisis de las variables numéricas en las cotizaciones B2C muestra un comportamiento altamente asimétrico y disperso. El valor total (valor) tiene una media de 36,9 pero un máximo superior a 366.000, evidenciando *outliers* significativos y un sesgo positivo. El 75% de los registros están por debajo de 27,6, lo que indica un mercado dominado por productos de bajo costo, aunque existen casos excepcionales de productos premium. La variable

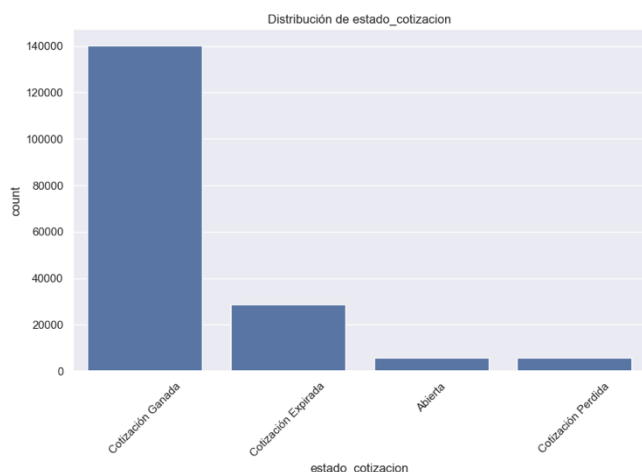


Ilustración 4 Histograma de distribución de estado de cotización

cantidad presenta una alta dispersión ( $std = 1.128$ ) y un valor máximo extremo de 438.000 unidades, lo cual también sugiere la existencia de registros atípicos o compras mayoristas aisladas.

En cuanto a las variables categóricas, específicamente “estado\_cotizacion”, se observa una clara predominancia de cotizaciones ganadas, que representan cerca del 77% del total. Le siguen en frecuencia las cotizaciones expiradas, mientras que las abiertas y perdidas tienen una participación

marginal y muy similar entre sí.

Este patrón sugiere un proceso comercial con alta efectividad de conversión desde la etapa de cotización, aunque también podría reflejar una sobre clasificación automática como "ganada" si no hay actualización posterior.

En cuanto a correlaciones, valor se relaciona débilmente con cantidad ( $r = 0.10$ ) y precio ( $r = 0.056$ ), lo cual es esperado ya que  $\text{valor} = \text{cantidad} * \text{precio}$ , pero las correlaciones bajas indican alta variabilidad en estas relaciones. El resto de las variables no muestra correlaciones relevantes con los datos económicos y operativos, y pueden considerarse identificadores sin valor predictivo. Este comportamiento refuerza la necesidad de limpieza de *outliers*, especialmente en cantidad, y de modelar considerando la alta varianza y fragmentación del mercado.

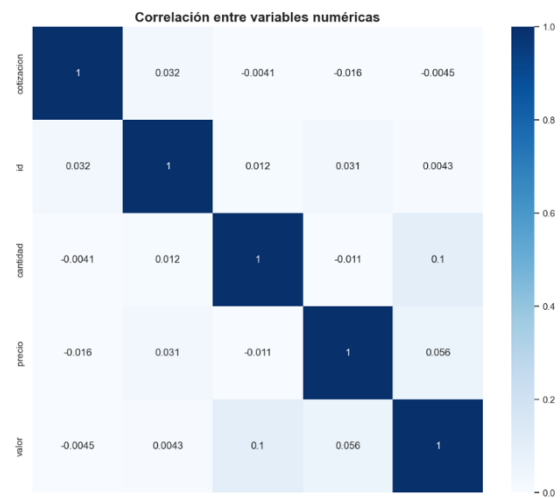


Ilustración 5 Correlación entre las variables numéricas del cotizaciones

Base Transaccional B2B  
(base 3 transaccional b2b)

Para comprender la dinámica comercial del segmento B2B, se realizó un análisis exploratorio detallado sobre la base de datos brindada, la cual contiene 25,866 transacciones con 10 variables explicativas. Esta base representa información transaccional relevante, destacando aspectos clave como el tipo de producto adquirido, su categoría y subcategoría, la fecha de compra, el valor total de la transacción y su alineación con el portafolio estratégico B2B. Es importante mencionar que se evaluó la calidad de los datos, destacando que no se identificaron valores nulos ni registros duplicados. Esto facilitó un análisis directo, sin necesidad de limpieza adicional. La variable “fecha\_factura” fue correctamente transformada al formato *datetime*, permitiendo futuros análisis temporales como identificación de estacionalidades o tendencias comerciales.

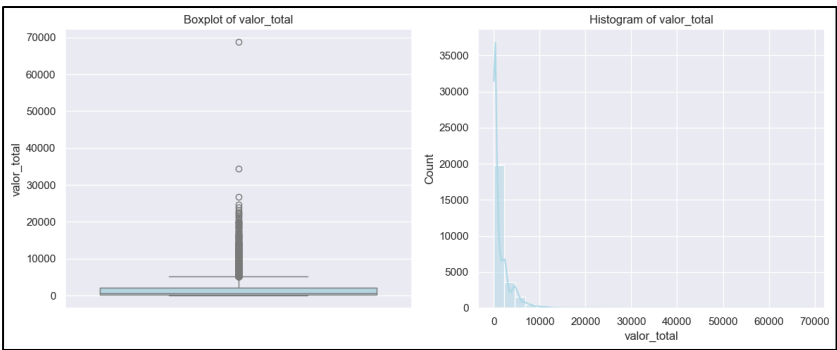


Ilustración 6 Histograma y boxplot de "valor\_total"

Respecto a las variables numéricas, la variable “valor\_total” mostró una distribución claramente sesgada hacia la derecha, con numerosos valores atípicos, como se evidencia en los histogramas y *boxplots* realizados. El 75% de las transacciones están por debajo de aproximadamente 2,200 unidades

monetarias, destacando una concentración de valores bajos a medios, mientras que algunos valores extremadamente altos alcanzan hasta cerca de 68,750 unidades monetarias. Estos valores atípicos podrían estar asociados a transacciones puntuales de alto valor,

posiblemente correspondientes a productos premium o adquisiciones de clientes corporativos clave.

La variable “alineación con portafolio estratégico B2B” mostró una distribución altamente concentrada alrededor de cero con una desviación estándar muy baja (0.000155), indicando que la mayoría de las transacciones están relativamente cercanas

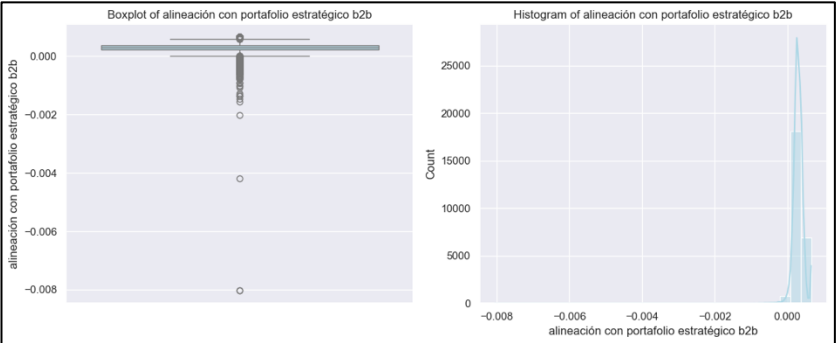


Ilustración 7 Histograma y boxplot de "alineación con portafolio estratégico B2B"

a un valor neutro. Sin embargo, la media (0.000294) y el valor máximo observado (0.000654) son tan bajos que, desde un punto de vista práctico, se interpretan como una baja alineación con el portafolio estratégico. Esto implica que muchas de las compras realizadas, aunque no necesariamente negativas, no parecen ser las más efectivas ni acordes con la estrategia definida por la empresa.

La presencia adicional de valores negativos, llegando hasta -0.008, enfatiza aún más la existencia de transacciones que claramente se alejan del enfoque estratégico deseado. En términos prácticos, esto representa una importante oportunidad para reevaluar y ajustar las estrategias comerciales, asegurando que las futuras transacciones prioricen productos con mayor valor estratégico para la compañía, optimizando así el uso de recursos y aumentando la efectividad comercial general. No obstante, se entiende que, al tratarse del segmento B2B, existe una limitación inherente al control directo sobre los productos que terceros comercializan. Por ello, se podría pensar en implementar campañas dirigidas específicamente a incentivar o patrocinar productos más alineados con la estrategia de la

empresa, motivando así a los aliados comerciales a promover aquellas referencias que mejor se ajustan a los objetivos estratégicos corporativos.

En el análisis de correlación, se identificó una correlación negativa débil ( $r = -0.13$ ) entre “valor\_total” y “alineación con portafolio estratégico b2b”. Esta asociación, aunque pequeña, indica una leve tendencia en que transacciones de mayor valor podrían estar menos alineadas con el portafolio estratégico. Esto podría significar que los clientes con transacciones más altas podrían estar

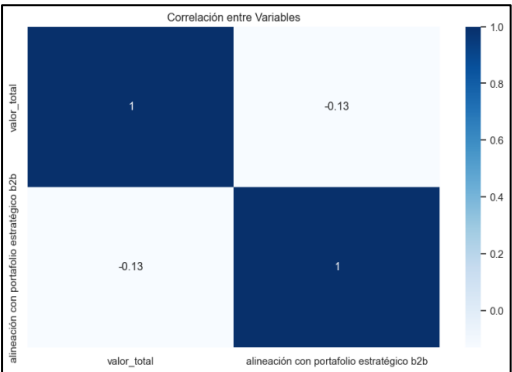


Ilustración 8 Gráfica de correlación de las variables numéricas B2B

adquiriendo productos menos prioritarios desde la perspectiva estratégica de la empresa. Tal hallazgo, aunque limitado en magnitud, ofrece información valiosa para futuras

revisiones estratégicas y comerciales que permitan aumentar la alineación estratégica de las transacciones más significativas en términos monetarios.

Desde el punto de vista categórico, se observa una concentración significativa en pocos actores. A nivel de clientes B2B, más del 50% de las transacciones están concentradas en el cliente B2B\_03, seguido por B2B\_02 y B2B\_01. Esto implica una fuerte dependencia comercial en unos pocos aliados estratégicos. Por municipios, destaca Fusagasugá con más del 62% de las transacciones, seguido por Villa de Leyva y Madrid. Un pequeño porcentaje (1.5%) tiene registros con municipio desconocido ("#"), lo cual debería revisarse si se desea realizar análisis geográficos más precisos. En cuanto a zonas, Cundinamarca representa el 71% de las ventas, seguida de Boyacá con un 27%. En cuanto a las categorías de productos, tanto a nivel macro como específico, la distribución también es bastante desigual. Las cinco principales categorías macro concentran más del 78% del total de transacciones, siendo la categoría "cat\_b2b\_macro\_1" la más destacada con el 22.3%. A nivel más granular, se identifican más de 130 subcategorías y más de 2,500 productos únicos, aunque nuevamente, una minoría de ellos acumula la mayoría del volumen de transacciones. Por ejemplo, el producto más vendido apenas representa el 1.11% del total.

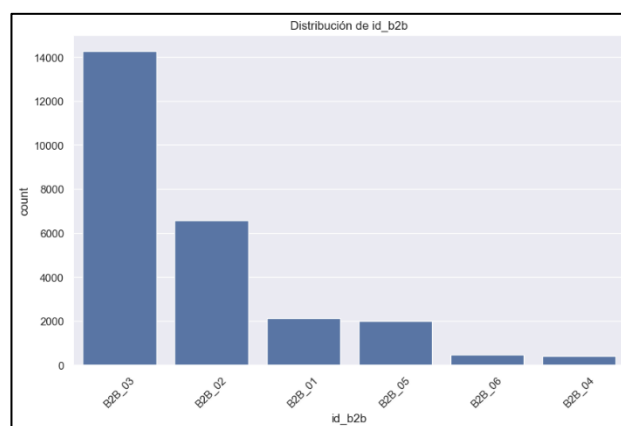


Ilustración 9 Distribución de los compradores B2B

Los hallazgos revelan una estructura de ventas altamente concentrada, tanto en un número reducido de clientes como en un subconjunto limitado del portafolio de productos, lo que resalta la necesidad de adoptar estrategias segmentadas más precisas. Una vía clave para abordar esta situación es diversificar el portafolio transaccional, impulsar la rotación de productos estratégicamente alineados y disminuir la dependencia comercial de un grupo tan acotado de clientes y referencias. No obstante, considerando que en el modelo B2B la empresa no ejerce control directo sobre las decisiones de compra de sus aliados, la implementación de un sistema de recomendación debe enfocarse en una lógica de acompañamiento inteligente. Aprovechando los datos detallados disponibles tanto del canal B2B como del B2C —incluyendo información geográfica como municipios y departamentos— es posible identificar patrones de consumo específicos por zona. Con base en esta evidencia, el sistema podrá sugerir al aliado comercial productos con alta demanda en su región, motivándolo a incorporar dichas referencias para responder mejor a las necesidades locales. Así, lejos de imponer decisiones, el modelo actúa como un facilitador estratégico que orienta al cliente hacia decisiones comerciales más informadas.

### Enriquecimiento de la base de datos

En el contexto de este reto, donde se busca desarrollar un modelo de recomendación inteligente capaz de personalizar la oferta de productos y mejorar la experiencia del cliente, se identificó la necesidad de complementar las bases de datos transaccionales con información adicional que permitiera captar mejor el contexto y las necesidades de los consumidores. Aunque inicialmente se evaluó la posibilidad de incorporar variables individuales como género, ocupación o preferencias personales, las bases disponibles — particularmente la B2C, que cuenta con el mayor nivel de granularidad— no incluyen este tipo de atributos. Si bien contiene datos como edad, municipio, punto de venta y producto adquirido, estos no son suficientes para capturar completamente los factores que inciden en la decisión de compra. Frente a esta limitación, se optó por enriquecer la información con variables externas de carácter demográfico, que pudieran contextualizar la demanda a nivel territorial. Se incorporaron así datos agregados por municipio y/o departamento, como porcentaje urbano, nivel de empleabilidad y otras condiciones socioeconómicas.

Este enfoque permite al modelo de recomendación ir más allá de los historiales de compra y considerar factores que inciden en las necesidades reales de los consumidores según su entorno. En lugar de imponer decisiones dentro del canal B2B —donde el control directo sobre lo que el tercero decide comprar es limitado—, el sistema podrá sugerir productos relevantes basados en comportamientos observados en la misma zona geográfica. Por ejemplo, si en determinado municipio los datos B2C muestran alta demanda de un producto específico, se puede recomendar al cliente B2B que lo incluya en su portafolio para satisfacer la demanda local. Con incluir esta información extra, se busca que el modelo no solo mejore la conversión de ventas y facilite el trabajo de los asesores comerciales, sino que también funcione como una herramienta estratégica para alinear las decisiones comerciales con los objetivos de negocio, maximizando así el valor de los datos y la eficiencia operativa.

#### Base Transaccional B2C (base 1 transaccional)

Con el fin de fortalecer el modelo de recomendación y capturar mejor el comportamiento del consumidor final, se decidió enriquecer la base transaccional del canal B2C a través de la construcción de nuevas variables derivadas directamente del historial de compras. Estas métricas permiten caracterizar a cada cliente desde múltiples dimensiones: intensidad de consumo (“total\_productos”, “total\_gasto”), sensibilidad al precio (“precio\_promedio”), frecuencia de compra (“num\_pedidos”), monto promedio por transacción (“ticket\_promedio”), profundidad por pedido (“cantidad\_promedio”) y diversidad en el portafolio consumido (“categorias\_diferentes”). En conjunto, estas variables no solo mejoran la granularidad del perfilamiento, sino que habilitan una segmentación más precisa para personalizar las recomendaciones según patrones de comportamiento reales.



Adicionalmente, se incorporaron variables externas provenientes de fuentes oficiales como el DANE, con el propósito de capturar el contexto demográfico y socioeconómico de los territorios en los que se realizan las compras. Se incluyeron atributos como la edad e ingreso laboral promedio, el porcentaje urbano, el índice de pobreza multidimensional (IPUG), el coeficiente de GINI y el total de edificaciones en obra por municipio o departamento. Estas variables permiten interpretar la demanda no solo desde la perspectiva individual, sino también desde las condiciones estructurales del entorno. Por ejemplo, el ingreso promedio y la urbanización pueden influir directamente en la capacidad adquisitiva y en la naturaleza de los productos más demandados, mientras que las edificaciones en obra ofrecen señales de crecimiento territorial que podrían anticipar futuras oportunidades de consumo.

Esta integración de información transaccional con variables contextuales responde a la necesidad de desarrollar un sistema de recomendación más robusto, que no se limite a repetir patrones de compra, sino que sea capaz de “entender” al cliente desde una lógica territorial, económica y social. Con ello, el modelo no solo personaliza de forma más efectiva, sino que también se convierte en una herramienta estratégica para la empresa, al alinear sus decisiones comerciales con el potencial real de cada zona y con el perfil específico de sus consumidores.

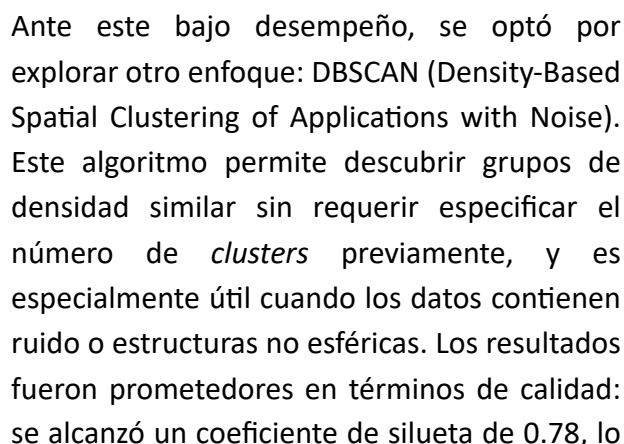
#### Base Transaccional B2B (base 3 transaccional b2b)

En el marco del desarrollo del modelo de recomendación, y reconociendo las particularidades del canal B2B —donde la compañía no tiene control directo sobre las decisiones de compra de sus aliados comerciales—, se incorporaron variables externas de carácter territorial con el fin de contextualizar mejor el entorno económico y urbano en el que estos operan. A diferencia del canal B2C, donde se cuenta con información granular por cliente, en el B2B el enriquecimiento se concentró en datos agregados a nivel municipal y departamental, lo que permite capturar dinámicas estructurales clave para orientar recomendaciones estratégicas.

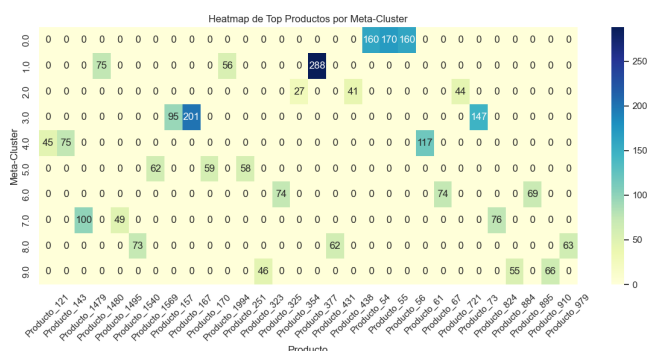
Se incluyeron variables como el total de edificaciones en obra, el promedio y la tasa de edificaciones por manzana (normalizadas por tamaño territorial), y la tasa por habitante, todas orientadas a medir el dinamismo urbano y el potencial de crecimiento físico del territorio. Estas métricas permiten identificar zonas en expansión o transformación urbana, donde es esperable una mayor demanda de productos para el hogar, la construcción o la remodelación, alineados con el portafolio de la compañía. Adicionalmente, se integraron variables económicas como el total de unidades económicas activas y la participación económica relativa del municipio dentro del total nacional. Estos indicadores permiten estimar la densidad empresarial y el peso económico de cada zona, facilitando la

La inclusión de estas variables transforma el modelo de recomendación en una herramienta propositiva y sensible al contexto territorial. Lejos de imponer productos, el sistema puede sugerir aquellos con mayor probabilidad de aceptación local, basándose en evidencia concreta del entorno donde opera cada aliado. Este enfoque mejora la pertinencia de las recomendaciones, fortalece la relación comercial y permite una alineación más efectiva con la demanda real, lo que puede traducirse en mayores niveles de rotación, mejor experiencia para el consumidor final y una mayor fidelización al canal.

En el canal B2B, se exploró la posibilidad de identificar patrones de comportamiento mediante técnicas de agrupamiento, con el objetivo de segmentar a los aliados comerciales según características transaccionales comunes. Inicialmente se aplicó el algoritmo KMeans, una técnica de *clustering* ampliamente utilizada que agrupa observaciones en función de su cercanía a centros de masa (centroides) optimizados. Sin embargo, al evaluar la calidad de la segmentación obtenida, se observó un coeficiente de silueta de apenas 0.13, lo que sugiere que las agrupaciones generadas no eran claras ni coherentes, es decir, los puntos dentro de los *clusters* estaban poco cohesionados y separados entre sí.



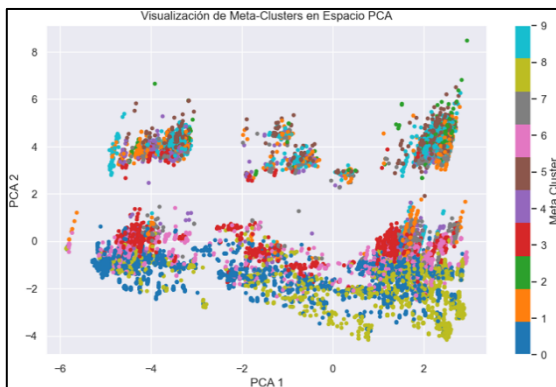
cual indica una segmentación mucho más definida. No obstante, DBSCAN generó un número excesivo de clusters —1,178 grupos— muchos de los cuales contenían muy pocos elementos (algunos con tan solo entre 2 y 10 observaciones), lo que dificultaba extraer conclusiones prácticas o



estrategias comerciales sólidas a partir de ellos.

Para superar esta limitación, se implementó una estrategia de realizar agrupación de los grupos encontrados. A partir de los *clusters* generados por DBSCAN (excluyendo los puntos identificados como *outliers*), se calcularon los centroides de cada grupo con base en sus variables numéricas. Posteriormente, se aplicó nuevamente KMeans, esta vez sobre estos centroides, con el fin de consolidar los *clusters* en grupos más interpretables y estratégicamente útiles. Esta técnica permitió sintetizar la gran cantidad de *clusters* originales en 10 "meta-clusters", agrupando a los aliados comerciales en subconjuntos de mayor tamaño y coherencia.

*Ilustración 11 Distribución de productos de los meta-clusters*



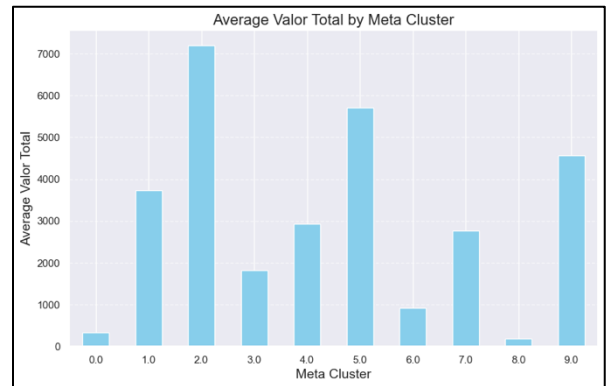
*Ilustración 12 Gráfica PCA de los clusters*

El resultado fue una estructura más manejable y valiosa desde el punto de vista analítico. lo que permite distinguir perfiles de clientes B2B con diferentes patrones de compra. A pesar de no contar con descripciones detalladas de los productos, se asumió que aquellos con numeraciones similares podrían pertenecer a una misma línea o tipo. Bajo esta premisa, se destacan tres meta-clusters que presentan patrones de compra claros y diferenciados:

- **Meta-Cluster 0 (8,416 registros):** Este es el grupo más numeroso del análisis y se caracteriza por una fuerte concentración en tres productos consecutivos: Producto\_55, Producto\_54 y Producto\_56, cada uno con más de 160 unidades vendidas. Esta homogeneidad sugiere un perfil de cliente altamente especializado, posiblemente vinculado a una sola línea del portafolio. Sin embargo, a pesar de su tamaño, este grupo presenta uno de los valores promedio de compra más bajos. Esto sugiere que, aunque los clientes realizan muchas compras, lo hacen sobre productos de bajo precio. En consecuencia, se trata de un grupo con alto volumen, pero bajo valor promedio, lo cual lo convierte en un segmento ideal para estrategias que busquen aumentar el valor por transacción.
- **Meta-Cluster 1 (1,533 registros):** Este grupo muestra una marcada preferencia por Producto\_377, con 288 unidades vendidas, muy por encima del resto. También destacan Producto\_1480 y Producto\_1994, aunque en menor proporción. Este patrón sugiere una alta dependencia de un producto clave, probablemente vinculado a una necesidad operativa específica. El valor promedio de compra en este grupo se encuentra en un rango medio, lo que indica que, además de ser consistentes en sus elecciones, los clientes realizan compras de valor moderado. Dado su comportamiento focalizado y estable, este meta-cluster representa un nicho estratégico que podría beneficiarse de

recomendaciones complementarias, programas de fidelización o condiciones comerciales especiales para fortalecer la relación y fomentar la retención.

- Meta-Cluster 3 (2,457 registros): Este grupo se distingue por una mayor diversidad de productos con alto volumen de ventas: Producto\_167 (201 unidades), Producto\_73 (147) y Producto\_157 (95). Este patrón sugiere un comportamiento de compra consistente y centrado en volumen, típico de revendedores o distribuidores en mercados con alta rotación. Sin embargo, a pesar de su tamaño y variedad, el grupo presenta uno de los valores promedio de compra más bajos. Esto indica que adquieren productos económicos en grandes cantidades, lo que refuerza su perfil de clientes mayoristas. Para este segmento, las estrategias comerciales deberían enfocarse en garantizar disponibilidad, eficiencia logística y estabilidad en precios.



*Ilustración 13 Valor total promedio de los clusters*

Cabe mencionar que también se intentó replicar el proceso de agrupamiento sobre la base transaccional B2C, con el fin de identificar segmentos de clientes individuales y sus patrones de consumo. No obstante, debido al tamaño y granularidad de dicha base —que contiene un volumen significativamente mayor de registros y variables codificadas—, el proceso resultó inviable desde el punto de vista computacional. Al aplicar técnicas como KMeans o DBSCAN sobre el conjunto completo, los modelos arrojaban errores relacionados con la dimensionalidad y la capacidad de procesamiento de la máquina, imposibilitando la convergencia o ejecución exitosa del algoritmo. Esto se debe a que, en *datasets* de alta dimensión, los algoritmos de *clustering* requieren recursos intensivos y pueden volverse ineficientes o inestables. Por esta razón, se priorizó el análisis de agrupamientos en el canal B2B, donde la estructura de datos era más manejable y permitía obtener resultados útiles y accionables en términos comerciales.

## **METODOLOGÍA IMPLEMENTADA**

Para seleccionar la metodología adecuada, partimos del reconocimiento de que se trata de un problema implícito, es decir, sin acceso directo a las preferencias declaradas por los clientes. Esto nos obligó a centrar los esfuerzos en comprender al usuario a través de su historial de compra y su contexto. Este enfoque se mantuvo a lo largo de todo el proceso, desde el enriquecimiento de datos hasta la elección de los algoritmos. Dada la alta dimensionalidad de las variables y el volumen considerable de registros, se concluyó que ningún modelo por sí solo sería capaz de capturar la complejidad del problema. Además,

debido a las limitaciones de recursos computacionales, fue necesario dividir el reto en subproblemas más manejables, evitando así la construcción de un modelo monolítico sobredimensionado y propenso al sobreajuste.

Con esto en mente se decidió implementar un algoritmo de LightFM, uno de clasificación supervisada XGBoost y finalmente un modelo de recomendación Híbrido que pudiera representar una combinación ponderada entre los algoritmos, para d alguna forma dar otra voz al problema. Esta solución se desarrolló para ambos contextos, con ciertas diferencias en su implementación, pero bajo la misma lógica.

**Algoritmo LightFM:** Se optó por este algoritmo ya que es un modelo de recomendación híbrido donde se combinan las ventajas del **filtrado colaborativo** (útil debido a la cantidad de clientes y compras realizadas) junto con el uso de la información de contenido (como las características de los productos).

- Marco Teórico

Es un algoritmo basado en **embeddings**, esto hace que las recomendaciones sean personalizadas, también con altas dimensionalidades. Se genera una función matricial que implementa un ranking. Para dicho ranking utilizamos la función WRAP, que penaliza los primeros lugares del ranking, útil en este contexto donde hay un grandísimo número de productos a ranquear. Se genera un producto punto entre el producto comprado con los demás vectores (otros productos) junto con el aprendizaje colaborativo:

$$\hat{r}_{ui} = \langle p_u, q_i \rangle + b_u + b_i$$

Donde,  $p_u$  es un vector de usuario,  $q_i$  es el vector de producto y  $b_u + b_i$  son los sesgos aprendidos.

- **Implementación (se dividió en tres etapas):**

*a. Construcción del Dataset*

Se mapearon los usuarios y los productos junto con sus características (debido a la alta dimensionalidad no fue posible usar todas las variables seleccionadas pero las que no se utilizaron en este modelo se usaron el en siguiente). Las variables seleccionadas fueron:

Modelo	Usuarios	Productos
B2C	Clúster, municipio, asesor, punto de venta, zona	Categoría macro, subcategoría, color
B2B	Municipio, zona, Total de unidades, Total de edificaciones en obra	Categoría macro, categoría, subcategoría

*b. Generación de Interacciones y Features*

Las interacciones se construyeron a partir del historial de compras. En el caso B2C, se utilizó una señal binaria (compra/no compra), mientras que en B2B se empleó el valor monetario total de compra como señal de intensidad. Se emplearon funciones personalizadas para construir listas de atributos por usuario y producto, que luego fueron vectorizadas mediante `build_user_features()` y `build_item_features()`.

### c. Entrenamiento del modelo

El modelo se entrenó utilizando el algoritmo WARP con una configuración estándar de `no_components=16` y `epochs=5`. Se utilizó `num_threads=4` para paralelizar el entrenamiento y mejorar el rendimiento. Para cada cliente, se utilizó el método `predict()` sobre todos los ítems no comprados, y se ordenaron por score descendente. Se devolvieron los *Top N* productos recomendados.

**Algoritmo XGBoost:** Se seleccionó este algoritmo ya que le daba una perspectiva distinta al modelo generando un modelo de aprendizaje directo a partir de atributos explícitos de los usuarios y los productos.

#### - Marco Teórico

XGBoost (*Extreme Gradient Boosting*) es un algoritmo de aprendizaje supervisado basado en árboles de decisión. Combina múltiples árboles secuenciales, donde cada árbol corrige los errores del anterior. Esto genera modelos precisos, robustos frente a *overfitting* (problema que ya estábamos experimentando con el anterior modelo) y eficientes en tareas de clasificación binaria, lo que buscábamos: predecir la probabilidad de que un cliente compre un producto. Este algoritmo optimiza una función objetivo que combina la pérdida logística con una penalización de regularización que controla la complejidad de los árboles:

$$\iota(\phi) = \sum_i \iota(y_i, \hat{y}_i) + \sum_k \Omega(f_k), \quad \Omega(f) = \gamma T + \frac{1}{2} \lambda |\omega|^2$$

Donde,  $\iota$  es la función de pérdida logística,  $\hat{y}_i$  es la predicción del modelo y  $\Omega(f_k)$  penaliza árboles con demasiadas hojas (T) o pesos grandes.

#### - Implementación

Se buscó implementar este clasificador binario que pudiera estimar la probabilidad de compra para cada cliente y cada producto, para esto se generaron también tres etapas:

### a. Construcción del dataset Tabular

Se generaron combinaciones cliente-producto con etiquetas 1-0 de compra y no compra, generando así valores negativos. Dichos valores se balancearon con los positivos.

### b. Enriquecimiento de variables

Para la base b2b fue necesario manejar las variables categóricas con un *Label Encoder* distinto a los *dummies* ya que estos aumentaban la dimensionalidad de manera exponencial. Finalmente, las variables seleccionadas fueron:

Modelo	Usuarios	Productos
B2C	edad, edad_promedio, ingreso_laboral_promedio, GINI, IPUG, cluster	Categoría macro, subcategoría, color, precio, alineación con portafolio estratégico
B2B	Municipio, zona, Total de unidades, Total de edificaciones en obra	Categoría macro, categoría, subcategoría

### c. Entrenamiento

Se utilizó XGBClassifier con hiperparámetros conservadores, priorizando interpretabilidad y estabilidad del modelo. El entrenamiento se realizó sobre un *split* estratificado de 80/20, separando entrenamiento y evaluación final (hold-out). También se aplicó validación cruzada en 5 pliegues para garantizar la generalización del modelo.

### Modelo de recomendación Híbrido

Teniendo la implementación de estos dos modelos complementarios, pero con limitaciones, ya que uno interpreta bien las preferencias implícitas (LFM) pero tiene dificultadas con la grande dimensionalidad, el otro balancea, pero tiene un alcance limitado de predicción. Por esto, se generó otro algoritmo de recomendación que combina las señales colaborativas y de contenido, ponderándolas para obtener un score final de recomendación más robusto.

Con la siguiente ecuación se busca establecer una combinación lineal entre ambos modelos,

$$score_{híbrido} = \alpha * score_{lfm} + (1 - \alpha) * score_{xgb}$$

Donde,  $\alpha$  es un hiperparámetro ajustable entre 0 y 1 que controla la influencia relativa de cada modelo,  $score_{lfm}$  es la predicción del modelo LightFM (colaborativo),  $score_{xgb}$  es la probabilidad estimada por el modelo XGBoost (contenido).

A partir de unos productos candidatos, se manejan las predicciones de ambos modelos y ya con ambos scores resultantes se ajusta el peso  $\alpha$  de la ecuación y se retorna el ranking. Dicho peso es ajustable ya que se considera que, a partir del análisis de los modelos, el equipo de analítica de Corona puede determinar el peso dependiendo cual modelo se ajusta de mejor manera al contexto y a las necesidades de la empresa retornando predicciones más precisas y coherentes.

Adicionalmente se realizó un demo subido en el siguiente enlace de Youtube: <https://youtu.be/MdN4GboogrM> donde se puede visualizar el modelo funcional y las predicciones que genera.

## MÉTRICAS E IMPACTO

Para medir la validez y la utilidad de la herramienta se calcularon las siguientes métricas específicas de rendimiento para cada modelo implementado.

AUC: Mide la capacidad del modelo para distinguir entre las clases, los productos que serán comprados y los que no.

Precisión: De los productos que el sistema predijo como “relevantes”, ¿cuántos realmente lo fueron? Esto se mide con esta fórmula:

$$\text{Precisión} = \frac{\text{Positivos}}{\text{Positivos} + \text{Falsos Positivos}}$$

Recall: De todos los productos que son relevantes, ¿cuántos se recomendaron?

$$\text{Recall} = \frac{\text{Positivos}}{\text{Positivos} + \text{Falsos Negativos}}$$

F1-Score: Es una métrica que pondera y combina las dos anteriores.

$$F1 = 2 * \frac{\text{Precisión} * \text{Recall}}{\text{Precisión} + \text{Recall}}$$

Las métricas de los algoritmos implementados con los datos generados fueron:

	B2B		B2C	
	LightFM	XGBoost	LightFM	XGBoost
Precisión	93.33%	80%	32.22%	84%
Recall	0.40%	81%	31%	86%
F1	0.80%	79%	31.60%	85%
AUC	86.79%	87.35%	92.86%	91.33%

Aunque XGBoost demostró un rendimiento superior en la mayoría de las métricas — especialmente en Recall (81%) y F1 (79%) en el canal B2B, y con resultados igualmente sólidos en B2C (Precision del 84% y Recall del 86%)—, LightFM aporta un valor complementario difícil de ignorar. Su enfoque basado en aprendizaje de representaciones



permite capturar similitudes complejas entre usuarios y productos, lo cual es esencial en entornos donde la personalización es prioritaria. Si bien sus métricas actuales son más limitadas, particularmente en cobertura, esto podría estar condicionado por las variables disponibles o el enfoque de modelado utilizado, lo que deja espacio abierto para mejoras mediante ajuste fino e integración con otras señales.

Los resultados reflejan comportamientos claramente diferenciados entre canales. En B2B, LightFM logró una precisión notable (93.33%), indicando una gran capacidad para recomendar productos acertadamente, aunque con un recall muy bajo (0.40%), lo que lo convierte en un modelo conservador. En contraste, XGBoost ofreció un equilibrio más robusto entre precisión y cobertura. En B2C, la dinámica se invierte: LightFM destaca por su excelente capacidad de ranking (AUC de 92.86%) pero con bajo desempeño práctico, mientras que XGBoost mantiene un rendimiento fuerte y balanceado, ideal para decisiones automáticas y recomendaciones operativas en puntos de venta o canales digitales.

Por ello, se justifica plenamente la elección de una arquitectura híbrida que combine ambos modelos. Esta integración permite capturar lo mejor de cada enfoque: la precisión fina y la capacidad de personalización de LightFM, con la amplitud predictiva y robustez de XGBoost. Ajustando el peso de cada modelo según el canal, el tipo de cliente o los objetivos comerciales, se obtiene una solución flexible, potente y alineada con las necesidades reales del negocio.

#### Impacto del modelo para Corona

La implementación del sistema de recomendación representa un hito estratégico para Corona, al consolidarse como una herramienta inteligente que transforma los datos transaccionales en decisiones comerciales altamente personalizadas, eficientes y alineadas con los objetivos corporativos. Este modelo no solo automatiza la sugerencia de productos relevantes, sino que lo hace con base en variables socioeconómicas, contextuales y de comportamiento, lo cual habilita una personalización avanzada tanto para clientes individuales del canal B2C como para aliados comerciales del canal B2B.

Desde el punto de vista operativo, el sistema contribuye directamente a optimizar los inventarios, incrementar la tasa de conversión, elevar el valor promedio por transacción y reducir el esfuerzo comercial necesario para identificar oportunidades de venta. En el canal B2C, por ejemplo, el modelo XGBoost alcanzó métricas de desempeño sobresalientes como una precisión del 84% y un AUC de 91.33%, lo cual valida su capacidad para anticipar las necesidades del consumidor final y sugerir productos adecuados en el momento justo. En el canal B2B, la precisión del modelo LightFM superó el 93%, aunque con menor recall, lo que permite pensar en combinaciones estratégicas entre modelos según el contexto y objetivos específicos (recomendación precisa vs. cobertura amplia).

Más allá del rendimiento técnico, el valor estratégico del modelo radica en su alineación con los objetivos corporativos de Corona: personalizar la experiencia, mejorar la eficiencia comercial, maximizar los ingresos y consolidarse como socio clave en el mercado nacional. El modelo de recomendación propuesto actúa como catalizador para esta visión, al transformar un portafolio amplio y complejo en propuestas específicas que simplifican la toma de decisión tanto para consumidores como para asesores y clientes institucionales.

Casos de referencia como Netflix demuestran el poder de este tipo de soluciones: su motor de recomendación le ahorra más de 1.000 millones de dólares anuales en retención de clientes y ha permitido alcanzar tasas de éxito superiores al 90% en la adopción de contenido original, muy por encima del promedio de la industria (Quin, 2024). Esta evidencia respalda el potencial de los sistemas de recomendación como palancas de innovación comercial y ventaja competitiva sostenible.

Asimismo, desde una perspectiva técnica, el enfoque híbrido adoptado en este proyecto — combinando filtrado colaborativo y aprendizaje supervisado— responde a las mejores prácticas globales en sistemas de recomendación. Tal como lo destaca Hsiao et al. (2025), la tendencia actual en empresas líderes es consolidar modelos fundacionales que centralicen el aprendizaje de preferencias y distribuyan conocimiento a múltiples aplicaciones, lo cual maximiza la eficiencia, escalabilidad y capacidad de adaptación de los sistemas. En el caso específico de Corona, este modelo se convierte en una solución lista para escalarse e integrarse a sus plataformas de ventas, cotización o CRM, permitiendo así cerrar el ciclo entre datos, analítica y acción. Se habilita así un nuevo nivel de inteligencia comercial, donde las decisiones de oferta y recomendación no son arbitrarias ni genéricas, sino impulsadas por evidencia concreta, tanto desde el historial del cliente como desde el contexto territorial.

Este tipo de soluciones no solo generan valor económico. También representan un avance en la forma en que Corona se posiciona frente al mercado: como una empresa que escucha, entiende y anticipa, que no solo vende productos, sino que construye experiencias y relaciones. Al facilitar al consumidor la elección del producto ideal —como lo busca la estrategia del reto—, se incrementa la satisfacción, se mejora la lealtad y se generan relaciones comerciales de largo plazo, más rentables y sostenibles.

## **RECOMENDACIONES Y CONCLUSIONES**

Los resultados obtenidos y la arquitectura desarrollada constituyen una base sólida para que Corona continúe avanzando en la personalización inteligente de su estrategia comercial. No obstante, para maximizar el impacto de esta solución en el mediano y largo plazo, se recomienda fortalecer tres frentes clave: capacidad computacional, calidad de datos y profundidad analítica.

En primer lugar, se requiere la disponibilidad de mayores recursos computacionales para entrenar modelos de mayor complejidad, explorar arquitecturas más sofisticadas y trabajar con un número creciente de variables y registros. Esto permitirá mejorar la precisión y escalabilidad del sistema de recomendación, especialmente en contextos de alta dimensionalidad como el canal B2C. Segundo, es fundamental que Corona consolide y amplíe sus fuentes de datos. Esto implica no solo enriquecer la información actual con nuevas variables contextuales (como ya se ha hecho con indicadores demográficos y económicos), sino también incentivar la captura de señales explícitas por parte del cliente, como calificaciones postventa, comportamiento digital en tiempo real o preferencias declaradas. La calidad, variedad y granularidad del dato es la materia prima que define la potencia del sistema.

Finalmente, se recomienda profundizar en el análisis de variables explicativas y su relación con decisiones de compra. Esto implica desarrollar nuevas métricas de comportamiento, segmentaciones dinámicas y experimentos controlados que permitan entender mejor qué factores influyen en la decisión de compra, tanto a nivel individual como territorial. Este tipo de analítica avanzada no solo mejora la recomendación, sino que se convierte en un insumo estratégico para múltiples áreas del negocio, desde marketing hasta desarrollo de producto.

### **PERSPECTIVAS FUTURAS DE NEGOCIO**

La implementación del sistema de recomendación en Corona no representa un punto final, sino más bien un punto de partida hacia una transformación más profunda en la manera en que la empresa se relaciona con sus consumidores, asesores y aliados estratégicos. Las perspectivas futuras se estructuran en tres dimensiones complementarias: técnica, comercial y estratégica.

**1. Perspectiva técnica: hacia una arquitectura evolutiva e inteligente:** Desde el plano técnico, se abren múltiples líneas de desarrollo. Una de las prioridades es el refinamiento continuo del modelo híbrido implementado. Esto implica optimizar el ajuste de hiperparámetros, incorporar nuevas funciones de pérdida adaptativas, e incluso explorar arquitecturas de metamodelos, que combinen los outputs de LightFM y XGBoost mediante redes neuronales o modelos de *boosting* adicionales. Siguiendo el ejemplo de empresas como Netflix, que ha evolucionado hacia un modelo fundacional centralizado capaz de transferir aprendizajes entre distintas tareas (Hsiao et al., 2025), Corona podría avanzar hacia una arquitectura unificada de recomendación capaz de adaptarse dinámicamente al canal, al contexto geográfico y al perfil del usuario.

Adicionalmente, la incorporación de aprendizaje continuo permitiría que el modelo se mantenga actualizado frente a cambios en el comportamiento de los usuarios, sin requerir reentrenamientos completos. Esto cobra especial relevancia en contextos como el B2B,

donde los ciclos de compra pueden ser estacionales y los patrones de demanda más sensibles a eventos externos como licitaciones, obras o decisiones gubernamentales.

**2. Perspectiva comercial: integración real-time y estrategias postventa:** Desde una visión de negocio, el sistema puede escalar más allá del canal analizado. Una evolución natural sería su integración en otros puntos de contacto como lugares de cotización web, canales de atención al cliente o aplicaciones móviles. Esto permitiría capturar información en tiempo real (clics, búsquedas, abandono de carrito) y retroalimentar el sistema con señales implícitas y explícitas para mejorar la precisión y personalización de las recomendaciones.

Otra línea clave es la implementación de estrategias postventa. Permitir que los clientes califiquen productos o experiencias abre la puerta a enriquecer los modelos con *feedback* explícito, habilitando la transición hacia modelos de aprendizaje reforzado o sistemas basados en satisfacción percibida. Empresas como Netflix han demostrado que redefinir sus métricas clave (de “vistas” a “minutos vistos”) permitió alinear mejor sus modelos con el éxito real de sus contenidos (O'Brien, 2024). De forma similar, Corona podría redefinir su métrica de conversión efectiva, incorporando no solo compra sino repetición, recomendación o satisfacción.

**3. Perspectiva estratégica: un motor de decisiones inteligente y transversal:** La visión de largo plazo posiciona este sistema como un motor de decisiones automatizado para múltiples áreas del negocio. Su alcance no se limita a ventas: puede guiar decisiones de surtido, campañas de marketing, gestión de inventario, expansión territorial o incluso desarrollo de producto. Por ejemplo, si el sistema detecta alta demanda de una categoría en una zona específica, puede sugerir ajustes en el abastecimiento o incluso en la estrategia de precios y promociones. En este sentido, el sistema trasciende su rol de recomendador y se convierte en un orquestador de valor.

Inspirándose en casos como el de Netflix, donde el sistema de recomendación no solo guía lo que el usuario ve, sino qué contenido se produce, cómo se lanza y dónde se promociona (Quin, 2024), Corona puede adoptar un enfoque similar en su ecosistema de productos y canales. La clave está en convertir los datos transaccionales y contextuales en inteligencia estratégica, no solo para personalizar experiencias, sino para anticipar movimientos del mercado y operar con ventaja.

Finalmente, como señala Ravindran (2023), los sistemas de recomendación efectivos no se construyen solo con buenos modelos, sino con claridad en los objetivos y en el impacto deseado. En el caso de Corona, el objetivo es claro: optimizar la oferta, personalizar la experiencia y maximizar ingresos. Pero el verdadero diferencial será lograr que este sistema evolucione al ritmo del negocio, incorporando cada nueva interacción como una oportunidad de aprender, adaptarse y servir mejor.

## Bibliografía

- Adhikari, S. (2019, 27 de febrero). *Building a Movie Recommendation Engine in Python using Scikit-Learn*. Recuperado de <https://medium.com/@sumanadhikari/building-a-movie-recommendation-engine-using-scikit-learn-8dbb11c5aa4b>
- Cámara de Comercio de Bogotá (s.f). *Mercado laboral*. Org.co. Recuperado de <https://www.ccb.org.co/informacion-especializada/observatorio/analisis-economico/mercado-laboral>
- Chandra, R. (2024, 3 de abril). *9 machine learning algorithms for recommendation engines*. Daffodilsw.com. Recuperado de <https://insights.daffodilsw.com/blog/machine-learning-algorithms-for-recommendation-engines>
- DANE. (s.f). *Empleo y desempleo*. Recuperado de <https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-y-desempleo>
- DANE. (s.f). *Gran Encuesta Integrada de Hogares - GEIH - 2024*. Recuperado de <https://microdatos.dane.gov.co/index.php/catalog/819/get-microdata>
- Datascientest. (2022, 30 de noviembre). *Machine Learning & Clustering: el algoritmo DBSCAN*. Formación en ciencia de datos. Recuperado de <https://datascientest.com/es/machine-learning-clustering-dbscan>
- Hsiao, K.J., Feng, Y., & Lamkhede, S. (2025, March 21). *Foundation Model for Personalized Recommendation*. Netflix Tech Blog. <https://netflixtechblog.com/foundation-model-for-personalized-recommendation-1a0bd8e02d39>
- Kavlakoglu, E. (2025, 10 de febrero). ¿Qué es el filtrado basado en contenido?. *Ibm.com*. Recuperado de <https://www.ibm.com/mx-es/think/topics/content-based-filtering>
- KMeans. (s.f). Scikit-Learn. Recuperado de <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>
- Longo, M. (2017, 8 de septiembre). *The What, Why and How of Recommendation Systems*. Medium. Recuperado de <https://medium.com/retargetly/the-what-why-and-how-of-recommendation-systems-810d98789f83>
- Nvidia. (s.f.). *Recommendation System*. Recuperado de <https://www.nvidia.com/en-us/glossary/recommendation-system>

- O'Brien, C. (2024, 8 de mayo). *The Unstoppable Success of Netflix*. Digital Marketing Institute. Recuperado de <https://digitalmarketinginstitute.com/blog/the-unstoppable-success-of-netflix>
- Ravindran, R. (2023, 1 de enero). *What are Recommendation Systems and How are Companies Using Them?* Recuperado de <https://rishika-ravindran.medium.com/what-are-recommendation-systems-and-how-are-companies-using-them-a5b08ff4df42>
- RPubs - Clustering Jerárquico en R. (s.f). Rpubs.com. Recuperado de <https://rpubs.com/mjimcua/clustering-jerarquico-en-r>
- Seeda, P. (2021, 31 de octubre). Towardsdatascience.com. Recuperado de <https://towardsdatascience.com/a-complete-guide-to-recommender-system-tutorial-with-sklearn-surprise-keras-recommender-5e52e8ceace1/>
- Sierra, L. F. (2024, 3 de mayo). *CENU 2024 - inicio*. Todo Lo Que Necesita Saber Sobre El Censo Económico | DANE. Recuperado de <https://censoeconomiconacionalurbano.dane.gov.co/>
- Quin, J. (2024, 25 de diciembre). *Netflix's Billion-Dollar Secret: How Recommendation Systems Fuel Revenue and Innovation*. Recuperado de <https://www.linkedin.com/pulse/netflixs-billion-dollar-secret-how-recommendation-systems-qin-phd-7zece/>
- Villalonga, R. (2017, diciembre 1). *Sistemas recomendadores híbridos*. Medium. Recuperado de <https://medium.com/@rvillalongar/sistemas-recomendadores-hibridos-93a6ff29500>