

Informe de hallazgos

Caso de estudio: comportamiento de clientes respecto a apertura de créditos de vivienda y libre inversión  

Por: Andrés Ethorimn

Motivación

Debido a la necesidad de generar estrategias encaminadas a mejorar el desempeño del área de Planeación Comercial y Ventas de ADL. El equipo de la dirección de Ciencia de Datos presenta su propuesta metodológica para encaminar las futuras campañas comerciales, de acuerdo con los resultados obtenidos tras nuestro estudio de los datos del comportamiento de nuestros clientes.

1. Objetivos

1. **Entendimiento de la apertura de productos:** analizar las características y en qué condiciones están los clientes abriendo cada tipo de producto: loan  y housing .
2. **Diseño de una metodología para recomendar el producto ideal para cada cliente:** las características de los clientes pueden ser heterogénea, y segmentar los clientes de acuerdo con características similares puede ser de suma utilidad para proponer estrategias útiles para futuras campañas comerciales del área de Planeación Comercial y Ventas.

2. Análisis exploratorio de los datos

La primera observación que hacemos de los datos suministrados es que existe un sesgo significativo hacia ciertos productos. Para observarlo, consideramos una codificación especial que tuviera presente los 4 posibles estados de apertura de productos, siendo estas:

- 0: el cliente no posee loan , ni housing 
- 1: el cliente posee loan , pero no housing 
- 2: el cliente posee housing , pero no loan 
- 3: el cliente posee tanto loan  como housing 

En este sentido, observamos que los estados más subrepresentados son 3 y 1, tal y como se muestra en la ilustración 1. Esta incidencia implica que la metodología más apropiada para garantizar una metodología que ofrezca resultados de calidad requirió de un **nuevo muestreo** de los datos de la muestra suministrada.

Realizamos un análisis ANOVA univariado con el fin de identificar cuáles de estas variables numéricas presentaban diferencias significativas entre los 4 posibles estados de apertura de productos. Destacamos que las características que más distinción ofrecen respecto a estas categorías fueron la edad del cliente, y sus ingresos y sus egresos.

Observamos que los clientes que adquirieron sólo un crédito de vivienda, o ambos productos fueron en promedio más jóvenes, y tenían más ingresos que aquellos que no adquirieron ningún

producto, o que sólo adquirieron créditos de libre inversión. Sin embargo, observamos este mismo comportamiento para el egreso, evidenciando una posible correlación entre ambas variables. La ilustración 2 muestra estos resultados

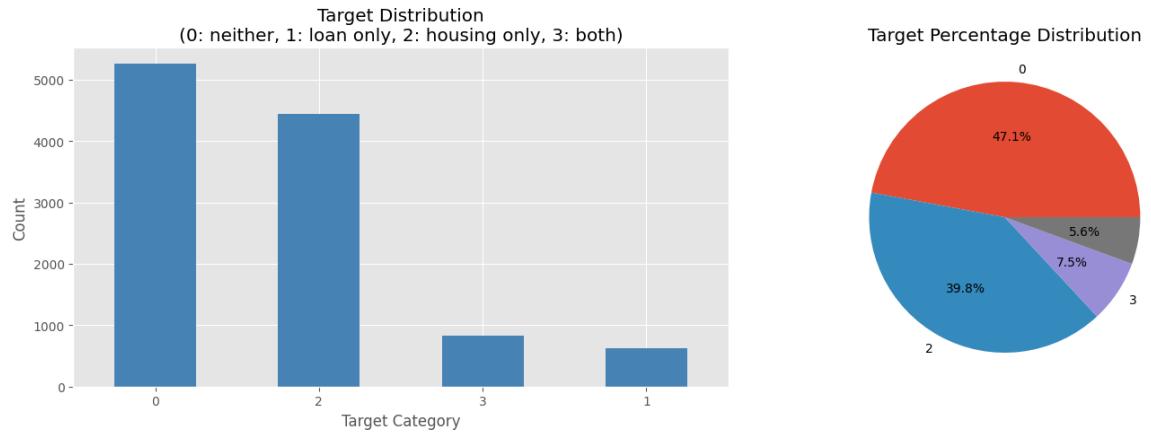


Ilustración 1: Conteo de datos pertenecientes a cada categoría de apertura de productos

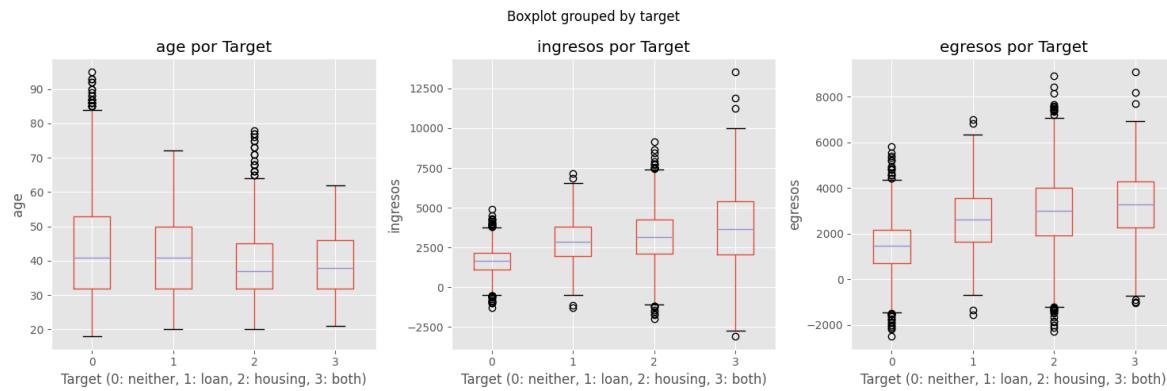


Ilustración 2: boxplots agrupados por categoría de apertura de productos de las variables con diferencias significativas entre categorías

Posteriormente, hicimos un análisis similar con la proporción de variables categóricas repartidas entre categorías de apertura de productos. Encontramos que con relación al empleo hay una marcada preferencia a que los clientes que no adquirieron ningún producto tengan una ocupación desconocida. También encontramos en este mismo grupo a los estudiantes, personas pensionadas, desempleadas y amas de casa. Por el contrario, observamos que las personas que adquirieron créditos de vivienda se encuentran empleados principalmente en el sector servicios, ejercen algún tipo de oficio, o tienen cargos administrativos y técnicos.

La ilustración 4 muestra el resultado de correlación de las diferentes variables numéricas respecto a las categorías de apertura. Observamos inmediatamente que los ingresos y los egresos son los que más correlaciones están con las categorías de apertura, como mostramos antes. Por otro lado, el balance y la edad muestran una leve relación inversa respecto a las categorías de apertura. Esto permite intuir que las personas más jóvenes y con menor liquidez presentan mayores problemas para adquirir nuestros productos.

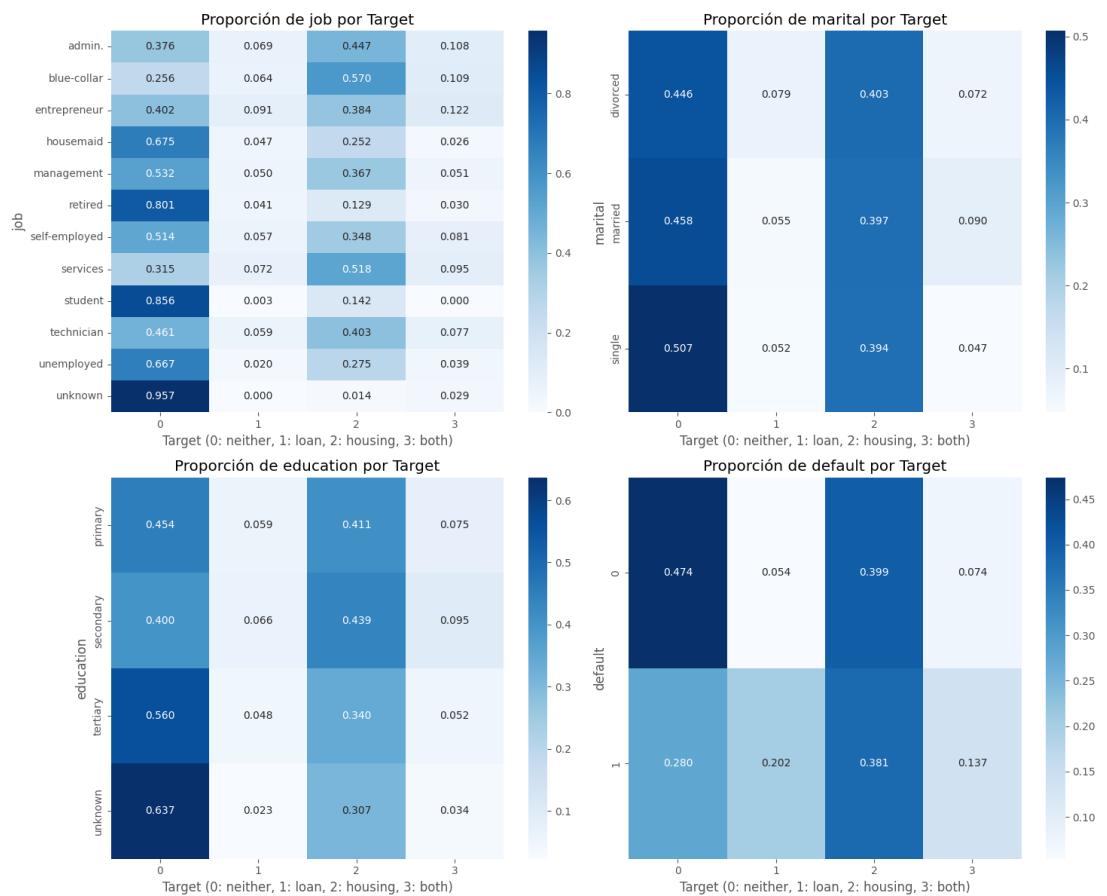


Ilustración 3: proporción de variables categóricas respecto a la apertura de productos

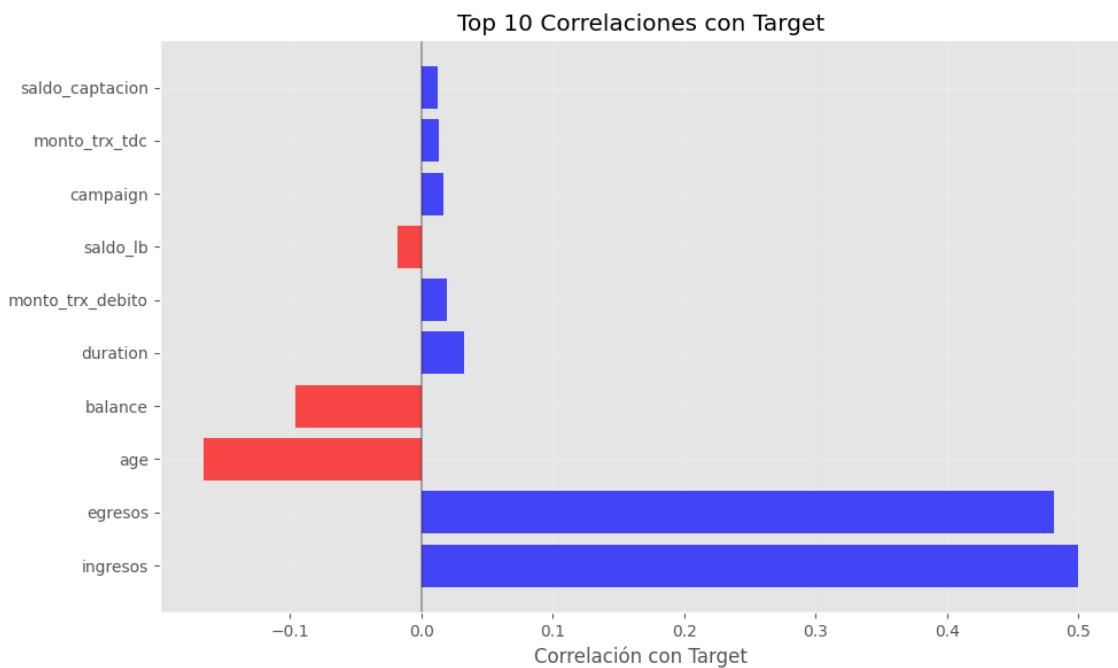


Ilustración 4: gráfico de correlación entre las variables numéricas y las categorías de adquisición de productos

3. Metodología de futuras campañas comerciales

Los datos de los clientes presentan características heterogéneas que vuelven necesario la creación de estrategias focalizadas para la maximización de la probabilidad de éxito al ofrecer nuevos productos. La mejor forma de llevarlo a cabo es mediante un método de segmentación que nos permite capturar grupos focales de clientes con características similares, y desarrollar campañas guiadas en estas.

En este sentido, propusimos una solución diseñando e implementando una metodología basada en el procesamiento de datos, el análisis y el aprendizaje no supervisado para a la creación de estos grupos focales. Por medio de estándares de desarrollo, proponemos la ejecución de este proyecto en cuatro fases:

1. *Capa de ingesta con estructuración de datos crudos. Carga en la capa <bronce>:*

En esta fase, la información puede provenir de diversas fuentes de datos estructurados o no estructura. Los datos que ingresen a los dominios digitales de ADL deben pasar por un proceso de ingesta en que se pueden almacenar directamente en un Datalake, o pueden pasar por un proceso de estructuración inicial antes de empezar su vida útil. Los datos útiles de recopilar en esta fase son información demográfica y financiera de los clientes, así como cifras históricas de interacciones previas como participación en campañas pasadas y la tenencia previa de productos. Para que este proceso sea eficiente, es necesario que la información tenga unos estándares mínimos de calidad para garantizar la gobernanza de datos. Entre estos estándares está la reducción de incidencia de valores nulos tanto como fuera posible, y el control del formato y la integridad de los datos.

2. *ETL con análisis exploratorio para la transformación, limpieza e imputación de valores faltantes donde fuese posible. Carga en la capa <plata>:*

En esta fase, los datos que fueron ingresados a través de la capa de ingesta son limpiados. Para esto podemos optar por eliminar los datos faltantes. Sin embargo, debido al desbalance de las categorías de adquisición de productos, así como a la reducida cantidad de datos hacen de esta estrategia peligrosa si no controlamos la información que vamos a perder. Por tanto, otra alternativa mejor es implementar métodos de imputación de datos. Para imputar los datos —que es darle valores a todos o una parte de los registros nulos, podemos usar una variedad de heurísticas de conformidad a la naturaleza de los valores no nulos de las variables a trasformar y su comportamiento. En este ejercicio, usamos métodos tan simples como la imputación de valores usando la mediana de la característica a completar, hasta métodos basados en modelos de aprendizaje automático auxiliares que permitieran imputar valores con base a los de las características adyacentes.

3. Última transformación de datos usando técnicas **reducción de dimensionalidad** como PCA en las características procesadas. Entrenamiento de modelos de clustering (agrupamiento) para la detección de grupos con características similares. Carga en la capa <oro>:

Cuando obtenemos nuestras **características trasformadas**, podemos usarlas para alimentar modelos de aprendizaje automático. En este caso, como tratamos de detectar patrones en los datos heterogéneos, lo mejor que podemos hacer es implementar aprendizaje no supervisado en forma de clustering (agrupación). Pero antes, podemos aplicar un proceso de aprendizaje intermedio llamado reducción de dimensionalidad, en que buscamos implementar técnicas reducir la complejidad de los datos resaltando las

características más importantes. El algoritmo PCA es una opción común y efectiva para conseguirlo, pues reducir el número de características mientras incrementa la varianza entre los datos transformados, preservando una gran parte de la información de los datos originales.

Después de transformar nuestros datos con PCA, podemos hacer experimentos con diversos algoritmos de clustering, en búsqueda de clasificar los datos en grupos independientes que nos permitan segmentar la información de manera estratégica. La forma apropiada de seleccionar uno entre varios métodos de clustering es realizar experimentos modificando los *índices* de estos modelos, buscando optimizar una serie de métricas de desempeño como el *silhouette score* o el *Calinski-Harabasz score*.

4. *Aplicar técnicas de análisis BI en estos grupos para detectar patrones en los datos que nos permitan diseñar mejores estrategias de planeación comercial y ventas.*

Finalmente, la segmentación final ofrecida por los diversos grupos o clusters serán útiles para segmentar los datos originales en grupos más pequeños y fáciles de manejar.

Idealmente, estos grupos tendrán ciertos patrones estratégicos que ayudarán a orientar mejores estrategias de negocio.

4. Resultados de la implementación de la metodología propuesta

Tras implementar la metodología propuesta, concluimos que podemos segmentar los clientes en 4 grupos focales con características que nos permiten entender mejor cómo orientar la próxima campaña de ventas.

La ilustración 5 muestra los análisis demográficos de los clientes clasificados en el **Cluster 1**. Al realizar a la respectiva comparación de los clientes en este grupo comparado con los otros, encontramos que concentra la mayoría de las personas jóvenes que poseen tanto créditos de vivienda  como créditos de libre inversión  . Adicionalmente, la ilustración 6 muestra datos financieros de este mismo grupo de personas.

Llama la atención cómo las personas en este grupo que poseen ambos productos poseen un mejor flujo de pagos, reportando mayores ingresos que gastos. Las personas que no poseen ningún producto en este grupo presentan el comportamiento contrario, tendiendo a tener más egresos que ingresos a medida que sus ganancias aumentan.

También observamos que en ambos subgrupos de apertura de productos, vemos que ambos tienen una alta prevalencia de educación terciaria, así como la tenencia de otros productos como vehículos o ahorros. Por tanto, una estrategia comercial que podría resultar interesante para este grupo de personas, que suelen ser las que mayor poder adquisitivo tienen para adquirir alguno de los productos o ambos, es ofrecer algún beneficio adicional que les represente un alivio en su flujo financiero. Esto es, opciones que les permitan aplacar la relación negativa entre ingresos y egresos, como planes de ahorro con tasas de interés favorables en caso de que adquieran alguno de los productos financieros suscritos; u opciones que les permitan mitigar gastos adicionales en los que estén inmersos, como compra de cartera o servicios para sus vehículos a precios exclusivos.

Demographic Analysis — Cluster 1

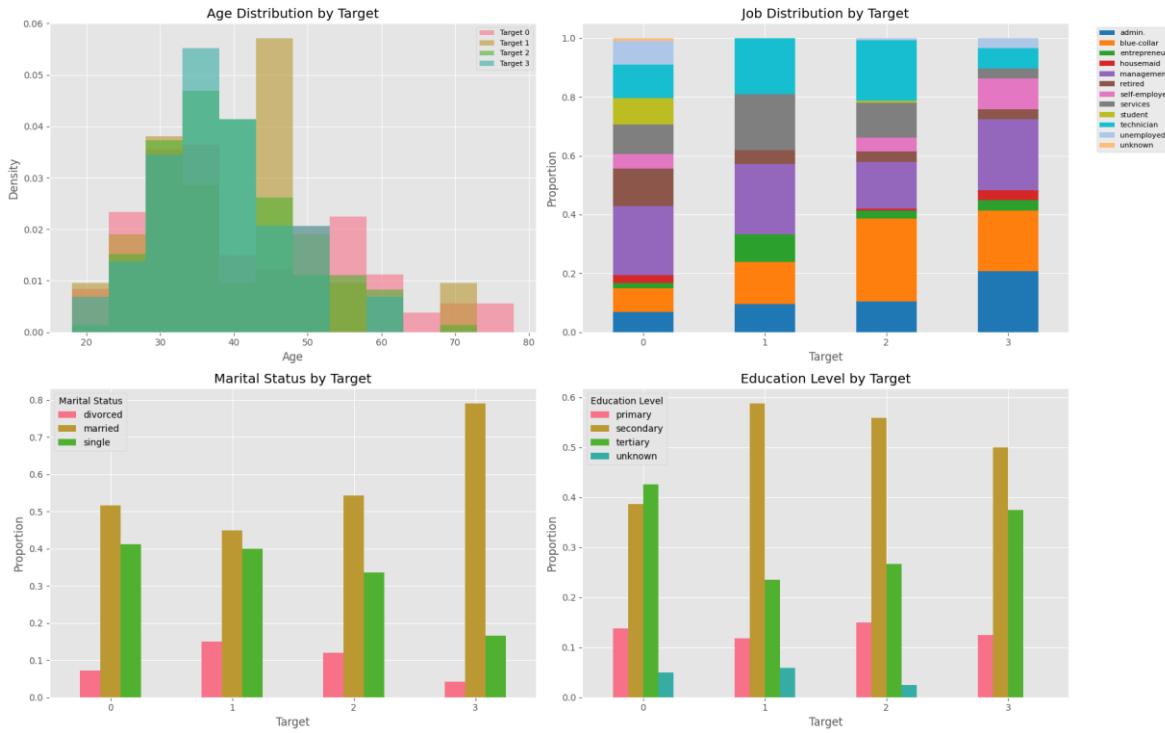


Ilustración 5: información demográfica de clientes que fueron clasificados en el "Cluster 1", de acuerdo a la metodología propuesta

Financial Behavior Analysis — Cluster 1

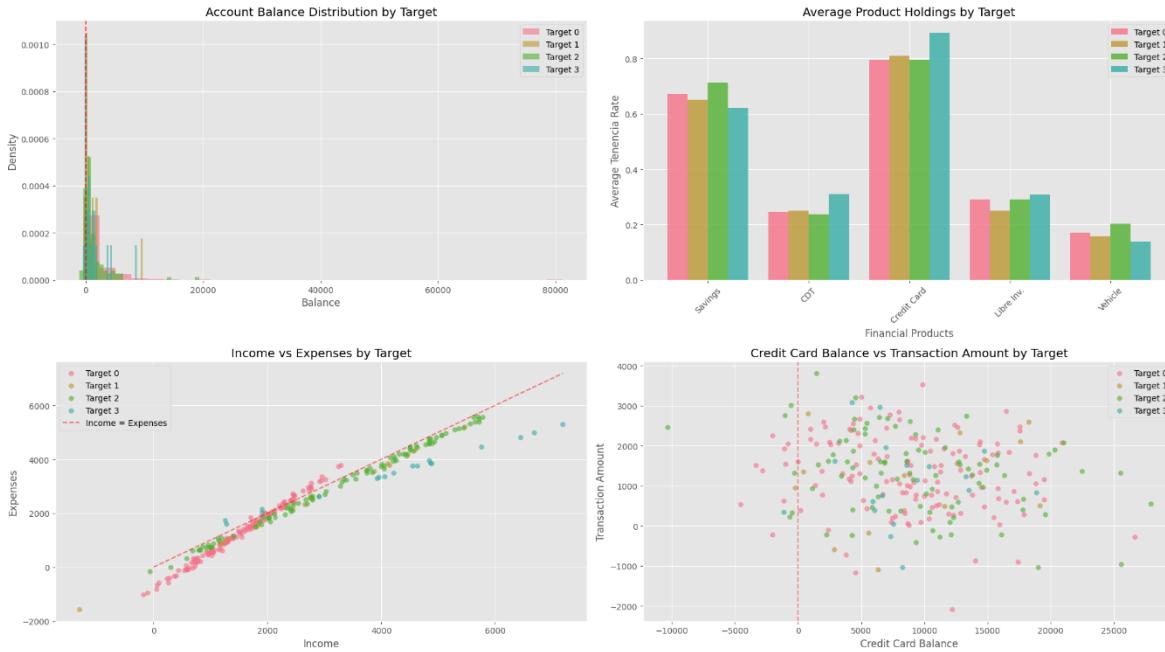


Ilustración 6: información financiera de clientes que fueron clasificados en el "Cluster 1", de acuerdo a la metodología propuesta

5. Próximos pasos y oportunidades de crecimiento con IA generativa

En general, las capacidades estratégicas que confiere la metodología propuesta son vastas, considerando que son 4 grupos focales subdivididos a su vez en 4 categorías de apertura de productos. A manera de resumen, podemos resumir las cualidades principales de los grupos focales de la siguiente manera:

- **Cluster 0:** clientes de mayor edad que la media con capacidad de inversión en créditos de vivienda y de libre inversión. Usualmente se encuentran ejerciendo algún oficio, tienen profesiones técnicas o son administrativos. Tienen menor capacidad de ahorro cuando sus ingresos son menores.
- **Cluster 1:** clientes más jóvenes que la media con capacidad de inversión en crédito de vivienda y de libre inversión. Usualmente se encuentran ejerciendo algún cargo gerencial. Suelen tener mayores ingresos y una capacidad de ahorro, pero no siempre es el caso cuando no adquieren ningún producto. Estudios superiores
- **Cluster 2:** clientes principalmente mayores y sin mucha capacidad de inversión en general. Hay una prevalencia de créditos de libre inversión por encima de los de vivienda. No se evidencia una capacidad de ahorro importante, pero tampoco un déficit.
- **Cluster 3:** clientes mucho más jóvenes que la media con capacidad de inversión en créditos de vivienda y de libre inversión. La mayoría están solteros. Estudios superiores.

Ahora bien, otra mejora en el flujo de trabajo del área de Planeación Comercial y Ventas de ADL es poder crear estrategias comerciales para cada subgrupo a mayor velocidad, para llegar a la mayor cantidad de potenciales clientes sin necesidad de perder precisión y capacidad persuasiva en el proceso.

Para ello, la IA generativa, y en particular, su moderna versión de Agentes IA son una estrategia poderosa para alcanzar esta meta. En primer lugar, la IA generativa tradicional ya pueden aportar ventajas al agilizar el proceso con la generación automatizada de visualizaciones y analíticas sencillos para la toma de decisiones. Sin embargo, algo más interesante viene cuando se combinan estas estrategias basadas en datos con agentes que sean capaces de tomar decisiones por sí mismos y automatizar flujos enteros de trabajo.

Un Agente IA puede integrarse con diversidad de entornos de desarrollo o empresariales. Esto permite automatizar algunas tareas repetitivas como la redacción de correos con reportes periódicas, así como la creación de contenido personalizado dadas las características y metadatos del cliente. Como funcionalidades auxiliares a las labores diarias del área de Planeación Comercial y Ventas están el agendamiento llamadas o videollamadas con clientes y proveedores. Estos Agentes IA pueden grabar y posteriormente analizar las conversaciones con los clientes para crear resúmenes ejecutivos sobre sus deseos o intereses.

En conclusión, la introducción de los Agentes IA a los flujos de trabajo de las empresas fue el último gran avance de las IA en entornos productivos. Este evento revolucionó las, hasta entonces, perspectivas respecto al potencial de la IA generativa clásica a convertirlo en una herramienta capaz de integrar muchas funciones para completar tareas cada vez más complejas. ADL tiene una maravillosa oportunidad de integrar estas tecnologías a sus flujos de trabajo, y así aprovechar todas las facultades que ofrece.