



Escuela de Administración

Modelo Predictivo de desistimiento aplicado a una constructora colombiana para proyectos de vivienda, un año después de la separación del inmueble.

Proyecto Empresarial – Trabajo de Grado

Presentado por:

Juan Felipe Vásquez, Yineth Juliana Velandia y Angie Carolina Rincón

Maestría en Business Analytics

Bogotá, D.C. 18 de junio de 2024



Universidad del
Rosario

Escuela de Administración
Escuela de Ingeniería, Ciencia y Tecnología

Maestría en Business Analytics

Modelo Predictivo de desistimiento aplicado a una constructora colombiana para proyectos de vivienda, un año después de la separación del inmueble.

Presentado por:

Juan Felipe Vásquez, Yineth Juliana Velandia y Angie Carolina Rincón

Bajo la dirección de:
Daniel Leonardo Cruz Castro

18 de junio de 2024

Tabla de contenido

Declaración de originalidad y autonomía	v
Declaración de exoneración de responsabilidad	vi
Lista de Figuras	vii
Lista de tablas	viii
Resumen ejecutivo	ix
Abstract	x
1. Introducción	3
2. Objetivo General	5
Objetivos Específicos	5
3. Alcance del Proyecto Aplicado	5
4. Metodología y Cronograma	7
5. Descripción de las Fuentes de Información	17
6. Entendimiento del Negocio	18
7. Descripción de la Problemática y Propuesta Resolutiva	22
8. Arquitectura Del Modelo De Datos	26
9. Análisis Exploratorio	27
10. Preparación De Datos	31
Construcción Del Pipeline	31
Perfilamiento de Datos	33
Ventanas de Tiempo LAG	40
11. Creación Del Modelo	43
Selección De Variables	43
Segmentación de Variables	45
Entrenamiento Balanceo	50
Modelo Árbol de Decisión	55
Despliegue Del Modelo	57
12. Estrategia de Gobierno de datos	59
13. Estimación y Administración de Riesgos	67
Identificación De Riesgos	67
Estrategias De Mitigación de Riesgos	72
14. Función De Utilidad (VAN) Valor Actual Neto.	76

15. Conclusiones E Insights	80
16. Recomendaciones Finales	82
17. Referencias	85

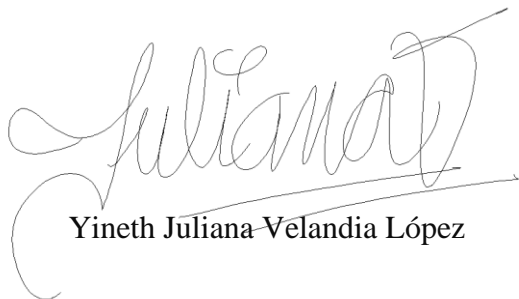
Declaración de originalidad y autonomía

Declaro(amos) bajo la gravedad del juramento, que he(mos) escrito el presente Proyecto Aplicado Empresarial (PAE), en la propuesta de solución a una problemática en el campo de conocimientos del programa de Maestría por mi(nuestra) propia cuenta y que, por lo tanto, su contenido es original.

Declaro(amos) que he(mos) indicado clara y precisamente todas las fuentes directas e indirectas de información y que este PAE no ha sido entregado a ninguna otra institución con fines de calificación o publicación



Angie Carolina Rincón Castiblanco



Yineth Juliana Velandia López



Juan Felipe Vásquez Peña

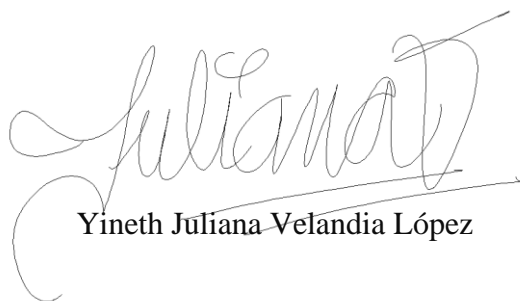
Firmado en Bogotá, D.C. el 18 de junio de 2024

Declaración de exoneración de responsabilidad

Declaro(amos) que la responsabilidad intelectual del presente trabajo es exclusivamente de su(s) autor(es). La Universidad del Rosario no se hace responsable de contenidos, opiniones o ideologías expresadas total o parcialmente en él.



Angie Carolina Rincón Castiblanco



Yineth Juliana Velandía López



Juan Felipe Vásquez Peña

Firmado en Bogotá, D.C. el 18 de junio de 2024

Lista de Figuras

Figura 1 Metodología CRISP-DM del proyecto	9
Figura 2 Tablero del diagnóstico inicial	21
Figura 3 Arquitectura del modelo analítico propuesto	26
Figura 4 Perfil del Desertor	29
Figura 5 Perfil del comprador.....	30
Figura 6 Pipeline Datafactory	32
Figura 7 Perfilamiento del dominio clientes.....	34
Figura 8 Perfilamiento de dominio de recaudo	35
Figura 9 Perfilamiento de dominio de ventas.....	36
Figura 10 Perfilamiento dominio de proyecto	37
Figura 11 Data Quality Scorecard	39
Figura 12 Ventanas Lags	42
Figura 13 Grafica de correlación de variables.....	44
Figura 14 Gráfica de correlación absoluta.....	45
Figura 15 Indicador Silueta de Algoritmos de Clustering.....	46
Figura 16 Porcentaje de Varianza de componentes del PCA	48
Figura 17 Gráfico PCA clusters.....	48
Figura 18 Curva ROC árbol de decisión	54
Figura 19 Matriz de confusión	55
Figura 20 Despliegue del modelo.....	58
Figura 21 Estrategia de Gobierno de datos.....	60
Figura 22 Nivel de madurez de gobierno de datos	63
Figura 23 Radar de nivel de madurez.....	64
Figura 24 Ruta de gobernabilidad	66
Figura 25 Niveles de riesgos	71

Lista de tablas

Tabla 1 Cronograma SCRUM del proyecto empresarial.....	10
Tabla 2 Diccionario de datos	27
Tabla 3 Variables ventanas de tiempo.....	40
Tabla 4 Comparación de Algoritmos de Clustering	47
Tabla 5 Identificación de clusters	49
Tabla 6 Comparativo de resultados de modelos	50
Tabla 7 Ventajas y desventajas de modelos evaluados	52
Tabla 8 Dataset de pruebas	56
Tabla 9 Propuesta de categorización de política gobierno de datos	62
Tabla 10 Modelo de gobernabilidad.....	65
Tabla 11 Identificación de riesgos de la constructora	68
Tabla 12 Estrategias de los riesgos.....	72
Tabla 13 Cálculo aproximado de costos iniciales	77
Tabla 14 Ingresos iniciales estimados	77

Resumen ejecutivo

El desistimiento de un cliente en el sector de la construcción es un fenómeno relevante que afecta tanto a los compradores como a las empresas constructoras. Se manifiesta cuando un comprador, previamente interesado y con una reserva mediante un pago inicial, decide cancelar su intención de compra antes de la entrega y escrituración del inmueble. Este comportamiento es recurrente en las operaciones comerciales de las constructoras en Colombia.

Para abordar este tema, el presente estudio se apoyó en una base de datos proporcionada por una empresa constructora, que contiene información histórica sobre los desistimientos de clientes. Esta base de datos se complementa con indicadores macroeconómicos del sector de la construcción. Se aplicó un proceso de Extracción, Transformación y Carga (ETL) para asegurar la calidad y pertinencia de los datos.

A través del análisis de los desistimientos, se identificaron factores clave que influyen en el abandono de la compra de inmuebles. Estos incluyen la pérdida de interés en el proyecto y el desempleo, entre otros.

Basándose en los hallazgos del estudio, se ha desarrollado un modelo predictivo basado en análisis de datos que evalúa y determina el riesgo de desistimiento para cada cliente. Este modelo incorpora variables clave como la edad, los ingresos, el sector laboral, el estado civil, el estrato socioeconómico y la ubicación geográfica del inmueble, así como indicadores macroeconómicos relevantes.

La implementación de este modelo permitirá a la empresa constructora tomar decisiones estratégicas fundamentadas, diseñar estrategias preventivas y correctivas, y

fortalecer su capacidad para mantener la estabilidad financiera y la satisfacción del cliente en un entorno altamente competitivo.

Abstract

This project aims to analyze customer withdraw in the context of the building company, a company in the construction sector, and propose strategies for improving customer retention. The study utilizes a database provided by the company, containing historical data on customer withdraw, and examines macroeconomic indicators in the construction sector for comparison. The data undergoes an Extract, Transform, and Load (ETL) process to ensure its quality and suitability for analysis.

Through an exploratory analysis of variables, key insights are gained regarding the factors influencing customer attrition. The findings highlight the significant impact of external factors and the economic climate on customer decisions.

Based on the analysis, recommendations are made to the building company, including continuous training for sales and business managers to enhance customer knowledge and retention skills. Consistent communication strategies are suggested to maintain customer interest in projects, while improvements in information systems are proposed to enhance data quality and decision-making processes.

Furthermore, a predictive model will be developed to assess the risk of customer attrition for each individual. This model will leverage the analyzed data and provide Prodesa with insights into customers who are more likely to discontinue their engagement. By identifying and prioritizing high-risk customers, Prodesa can implement targeted retention strategies and reduce overall withdrawal rates.

The conclusions emphasize the importance of understanding external influences, implementing targeted strategies, and utilizing a predictive model to manage customer attrition effectively. The project provides valuable insights and practical recommendations that can guide Prodesa in developing data-driven retention strategies and improving overall business performance.

1. Introducción

En el contexto socioeconómico actual en Colombia, las constructoras enfrentan desafíos significativos para mantener la rentabilidad y el crecimiento sostenible en un mercado altamente competitivo. Según la revista Forbes, la reactivación económica de 2021 tuvo al sector constructor como protagonista indiscutible. Lo anterior se explica por el incremento de 11,6% en el valor agregado del sector edificador y el inicio de obras nuevas equivalentes a más de 16 millones de metros cuadrados, según datos de la Cámara Colombiana de la Construcción (Camacol). Con datos como estos, aún es difícil entender lo que pasó tan solo dos años después, cuando el PIB de la construcción cayó un 8% anual con corte al tercer trimestre, siendo el sector con el mayor retroceso de toda la economía. Indiscutiblemente una de las amenazas más significativas para su estabilidad financiera es el desistimiento de clientes, un fenómeno que genera impactos negativos en la viabilidad de proyectos y la reputación de la empresa. El impacto de las tasas de interés, el cambio en la política de subsidios y el crecimiento de los costos de la construcción, más concretamente en los insumos, reflejaron incrementos exponenciales en los desistimientos, que sumaron 29.468 en 2023, y en una reducción del mercado de viviendas de interés social (VIS) cercana al 50 % al cierre del año pasado. Así es como el sector construcción pone manos a la obra para salir adelante en 2024. (Montes, 2024)

El desistimiento de clientes ocurre cuando los compradores potenciales, que inicialmente han abonado un saldo inicial en un proyecto inmobiliario desarrollado por una constructora, cancelan su intención de compra antes de la escrituración del inmueble. (Lesmes, 2023) Este fenómeno está relacionado con una serie de factores, como problemas económicos, cambios en las condiciones personales o incertidumbre en el mercado inmobiliario.

Para abordar esta problemática, se propone la creación de un modelo predictivo de clasificación de riesgo para calificar clientes en una constructora. El objetivo de este proyecto es desarrollar una herramienta basada en algoritmos que permita identificar temprano a los clientes con mayor probabilidad de desistir de su intención de compra. (Northcott, 2017)

Al emplear técnicas de análisis de datos y minería de información, es posible examinar variables relevantes como el perfil socioeconómico de los clientes, su historial crediticio, la ubicación del proyecto y las condiciones económicas del país, entre otros aspectos relevantes. El modelo resultante tendrá la capacidad de generar alertas y clasificar a los clientes según su nivel de riesgo de desistimiento, suministrando a la constructora información valiosa para la toma de decisiones estratégicas y el diseño de acciones preventivas y correctivas adecuadas.

La implementación de este modelo predictivo de indicadores de riesgo para el desistimiento de clientes brinda a la constructora una ventaja competitiva significativa al permitirle anticiparse a posibles cancelaciones de compra y tomar medidas proactivas para mitigar los riesgos asociados. Además, al reducir la tasa de desistimiento, se podrán optimizar los recursos financieros y operativos, mejorar la eficiencia y la rentabilidad, y fortalecer la relación con los clientes (Porter, 2015).

2. Objetivo General

Elaborar un Modelo Predictivo de desistimiento adaptado a una constructora colombiana centrada en proyectos de vivienda proyectos de vivienda de Interés Social (VIS), con el objetivo de identificar y comprender los factores determinantes que inciden en la decisión de los clientes de abandonar el proceso de compra, un año después de su separación o iniciación del negocio, proporcionando estrategias de gobierno de datos para obtener datos fiables.

Objetivos Específicos

- Recopilar y analizar datos de la constructora para identificar patrones, tendencias y relaciones relevantes relacionados con el desistimiento de compra de inmuebles en la organización.
- Clasificar a los clientes de la constructora en diferentes niveles de riesgo de desistimiento.
- Diseñar estrategias de gobernabilidad de datos para garantizar la calidad, confiabilidad y consistentes para realizar el modelo predictivo.

3. Alcance del Proyecto Aplicado

Se proyecta que el modelo se implemente para los clientes de la empresa en el año 2024; El primer paso de esta iniciativa será la recopilación de los datos relevantes necesarios para el desarrollo del modelo predictivo de la constructora. Esta fase comprenderá la recolección de información sobre los clientes, los proyectos de construcción en curso, las características demográficas, el historial de pagos, las interacciones con la empresa, así como cualquier otro dato considerado pertinente.

Una vez recopilados, los datos serán sometidos a un exhaustivo proceso de limpieza para asegurar su calidad y consistencia. Posteriormente, se aplicará el proceso de Extracción, Transformación y Carga (ETL) para preparar los datos de manera adecuada para su análisis.

Como una medida clave para garantizar la efectividad continua del modelo, se propone que este sea ejecutado de forma periódica cada 6 meses. De esta manera, se asegurará que el modelo esté alimentado con la información más reciente correspondiente a los 6 meses previos a su implementación inicial.

Se realizará un análisis en profundidad de los datos recopilados específicos de constructora para identificar patrones, tendencias y relaciones relevantes. Esto ayudará a comprender mejor los factores que pueden influir en el desistimiento de los clientes de la constructora y determinar las variables más relevantes para el modelo predictivo.

El modelo será capaz de clasificar a los clientes de la constructora en diferentes niveles de riesgo de desistimiento, utilizando las variables más relevantes identificadas en el análisis exploratorio de datos. Se emplearán algoritmos adecuados y se ajustarán los parámetros para obtener la mejor precisión posible.

La intención principal del proyecto es indagar, concretar y entender los datos de desistimiento de la constructora, por medio de las diferentes metodologías y procesos mencionados en el presente documento, para que la empresa pueda aumentar sus beneficios y pueda gestionar la información de ventas y desistimientos de la mejor manera, tomando decisiones basadas en datos.

En el marco del proyecto, se establecerá un proceso periódico donde el modelo de predicción generará informes mensuales identificando a los clientes con mayor probabilidad de

desistir de los proyectos en curso. Estos informes serán presentados en un formato digital accesible para el equipo de ventas.

Cada informe contendrá una lista de clientes identificados, junto con su nivel de probabilidad de desistimiento por proyecto. Además, se incluirá un análisis de los factores que contribuyen a esta probabilidad, ofreciendo insights sobre posibles preocupaciones de los clientes.

El equipo de ventas revisará estos informes mensuales y priorizará el contacto con los clientes identificados junto con los agentes de fidelización. La acción consistirá en discutir las preocupaciones identificadas y ofrecer soluciones personalizadas para abordarlas.

El resultado esperado de esta intervención temprana es reducir la tasa de desistimiento, lo que contribuirá a mejorar la satisfacción del cliente. Además, al abordar proactivamente las preocupaciones de los clientes, se espera fomentar relaciones más sólidas y duraderas con ellos, lo que a su vez puede traducirse en beneficios a largo plazo para la constructora.

4. Metodología y Cronograma

El proyecto de predictivo de indicadores de riesgo para el desistimiento de clientes en el sector de la construcción en Colombia se benefició de la metodología Scrum con sprints debido a varias razones.

El enfoque iterativo e incremental de Scrum permitió abordar el proyecto en ciclos cortos y repetitivos, conocidos como sprints. A través de los sprints, se pudieron establecer objetivos claros y alcanzables, lo que permitió al equipo priorizar las tareas y trabajar de manera iterativa en la creación del modelo predictivo.

La adaptabilidad y flexibilidad inherentes a Scrum (Marcelo Morandini a, 2021) resultan beneficiosas en proyectos de analítica, donde los requisitos y las necesidades pueden cambiar

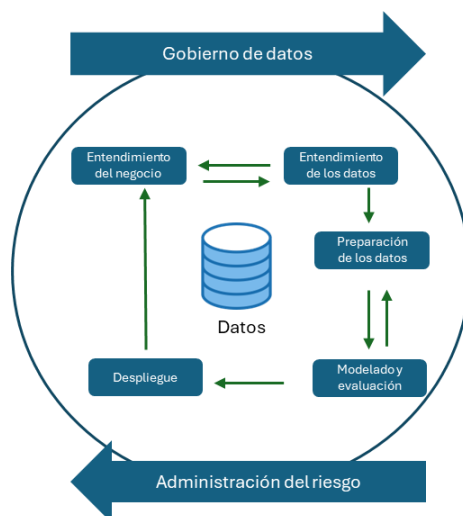
conforme se obtiene más información o se realizan descubrimientos durante el desarrollo. Scrum permite ajustar y adaptar los objetivos y prioridades del proyecto en cada sprint, facilitando la incorporación de cambios y garantizando una mayor alineación con las necesidades del negocio. Esto posibilita una mayor agilidad y capacidad de respuesta a medida que se adquiere una comprensión más profunda del problema y de las necesidades de los clientes

Además, la metodología Scrum fomenta la colaboración estrecha entre los miembros del equipo y los stakeholders. En proyectos de analítica, es crucial contar con la participación y el aporte de expertos, especialistas en datos y usuarios finales. Scrum promueve la comunicación constante, las reuniones de planificación y revisión de sprint, lo que facilita la colaboración y asegura que todos los involucrados estén alineados y comprometidos con los objetivos del proyecto (Ridewaan Hanslo, 2020). La interacción y la retroalimentación frecuentes entre los miembros del equipo y los stakeholders permiten ajustar rápidamente la dirección del proyecto y garantizar su éxito.

Asimismo, el proyecto combinó la metodología Ágil con el marco teórico de CRISP-DM (Cross-Industry Standard Process for Data Mining) para proporcionar a la constructora una solución analítica integral, este marco teórico fue enriquecido con el marco de gobernabilidad como se muestra en la Figura 1.

Figura 1

Metodología CRISP-DM del proyecto



El uso de CRISP-DM implicó seguir un proceso estructurado en seis fases: entendimiento del negocio, entendimiento de los datos, preparación de los datos, modelado y evaluación y despliegue. Este enfoque sistemático permitió comprender los objetivos de la constructora y las necesidades del negocio, recopilar y preparar los datos adecuadamente, desarrollar y evaluar el modelo predictivo de desistimiento y desplegarlo eficazmente en la compañía. Como resultado, la constructora se benefició al recibir una solución que cumplió con sus expectativas reales y que podía ser aplicada de manera práctica para reducir la tasa de desistimiento de clientes, esta ruta se detalla en cronograma visualizado en la Tabla 1.

Tabla 1*Cronograma SCRUM del proyecto empresarial*

	Dedicación en Horas	may- 23	jun- 23	oct- 23	nov- 23	mar- 24	abr- 24	may- 24
Sprint 1 - Entendimiento del Negocio								
1.1. Entrevista con la constructora	2	■						
1.2. Definición de objetivos y Alcance	12	■						
1.3. Recopilación de las fuentes de información	8		■					
1.4. Entendimiento del negocio - Diagnóstico Inicial	8		■					
Sprint 2 - Entendimiento de los datos								
2.1. Arquitectura del modelo	16		■					
2.2. Análisis exploratorio de los datos	16		■					
Sprint 3 - Preparación de los datos								
3.1. Construcción del pipeline	48		■					
3.2. Perfilamiento de los datos	12			■				
3.4. Calidad y Limpieza de los datos	24			■				
3.5. Ventanas de tiempo	8			■				
Sprint 4 - Modelado								
4.1. Selección y segmentación de Variables	8				■			
4.2. Balanceo de los 11 modelos	48					■		
4.3. Ejecución del modelo de árbol de decisión.	48					■		
4.4. Despliegue del modelo	16					■	■	
Sprint 5 - Evaluación y conclusiones								
5.1. Planteamiento de estrategias de gobernabilidad	24						■	
5.2. Planteamiento de administración del Riesgo	16						■	
5.3. Conclusiones y Recomendaciones	14						■	■

Un aspecto clave de Scrum fue la entrega temprana de valor, lo que permitió al equipo proporcionar incrementos de valor funcionales en cada sprint. De esta manera, se ajustaron las métricas, desarrollos y objetivos en línea con el modelo predictivo de riesgo de la constructora.

A continuación, se detallan las actividades, herramientas y entregables ejecutados en cada sprint.

Sprint 1 - Entendimiento del Negocio (mayo 2023 - junio 2023)

Entrevista con la constructora

Actividades. Reunión de planificación para comprender los objetivos del proyecto, entender los procesos claves de desistimiento y establecer los criterios para evaluar el éxito del modelo predictivo.

Herramientas. Planner para la gestión de tareas, Teams como plataforma de reunión con el equipo.

Entregable. Definición entre los equipos sobre los procesos claves del desistimiento de vivienda y las expectativas del proyecto.

Definición de objetivos y Alcance

Actividades. Se detallo los objetivos del proyecto y el alcance específico del análisis de datos y desarrollo del modelo.

Entregable. Definición de los objetivos, Alcance, Metodología y cronograma del proyecto con el equipo de la constructora.

Recopilación de las Fuentes de información

Actividad. Identificación y obtención de los conjuntos de datos relevantes, incluyendo información de los clientes, historial de recaudo, ubicación del proyecto.

Entregable. Definición de las fuentes de información a utilizar para el modelo analítico y para la estrategia de Gobernabilidad de los datos.

Entendimiento del Negocio

Actividades. Se realizó el levantamiento de las reglas del negocio, comprendiendo los requerimientos y planteando la estrategia para abordar la solución analítica. En esta etapa la constructora proporcionó una base de datos con los desistidos de la compañía, con esta base se hizo el primer diagnóstico para entender la situación inicial.

Herramientas. Power BI para la visualización del diagnóstico inicial con la base proporcionada por la compañía, Teams como plataforma para reunir al equipo.

Entregables. Tablero del Diagnóstico inicial del entendimiento del negocio.

Sprint 2 - Preparación de los datos (junio 2023)

En este segundo sprint, el equipo se centró en organizar, centralizar, normalizar y analizar los datos enviados por la constructora para determinar las variables más importantes para el modelo predictivo. Las actividades fueron:

Arquitectura del modelo

Actividades. se realizó la identificación del origen de los datos no estructurados por parte de la constructora, detallando el flujo de datos desde la captura de entrada hasta las tablas principales que componen el modelo analítico.

Entregables. Flujo de datos describiendo el proceso que se le realiza a los datos de la constructora.

Análisis Exploratorio de Datos

Actividades. Se llevó a cabo el análisis detallado de los conjuntos de datos, identificando patrones y tendencias. Se realizaron técnicas de visualización y estadística descriptiva para comprender mejor los datos proporcionados por la constructora.

Herramientas. SQL Server para análisis descriptivo de los datos, Power BI para la visualización del perfil del desistido y comprador.

Entregables. Diccionario de datos detallando los campos recolectados y normalizados de la constructora, Tableros descriptivos del perfil del desistido y comprador.

Sprint 3 – Preparación de los datos (junio2023 - octubre 2023)

En esta fase se ejecutó tareas de diagnóstico de calidad, estandarización, normalización y limpieza de los datos, los cuales fueron tratados desde Synapse de Azure, estableciendo una arquitectura de entidad - relación para la integridad de la información de la constructora.

Construcción del Pipeline

Actividades. Se realizó la extracción, Carga y transformación de los datos con el objetivo de normalizar y estandarizar los campos enviados por la constructora para realizar el perfilamiento y calidad de los datos, con el objetivo de seleccionar las variables con mejor calidad en el Dataset.

Herramientas. Data Factory para realizar el ETL (pipeline), Synapse para estructurar los datos en una zona cruda, transformarlos y disponerlos en tablas normalizadas.

Entregables. Tablas normalizadas y estandarizadas con la información histórica de los clientes, desistidos, ventas, proyectos y recaudo.

Perfilamiento de los datos

Actividades. Se ejecutó el perfilamiento de los datos para obtener una visión completa de su estructura, contenido y calidad. Esto incluyó la identificación de distribuciones, valores nulos, patrones, relaciones y otros aspectos relevantes de los datos.

Herramientas. SQL para manipulación del perfilamiento, Power BI para visualización del perfilamiento de los datos divididos en Dominios de información.

Entregables. Diagnóstico del perfilamiento de todos los datos entregados por parte de la constructora dividida en dominios de información.

Calidad y Limpieza de los datos

Actividades. Diagnosticar y mejorar la calidad de los datos, incluyendo la detección y corrección de valores atípicos y datos faltantes. Adicional realizó tareas de estandarización.

Herramientas. SQL para limpieza de valores atípicos, datos faltantes y estandarización de los datos, Power BI para visualización del diagnóstico de calidad dividido en Dominios de información.

Entregables. Datos otorgados por la constructora divididos en dominios de información con indicadores altos de calidad.

Sprint 4 - Modelado (noviembre 2023 - marzo 2024)

Selección de Variables

Actividades. El equipo utilizó la técnica de selección de variables de análisis de correlación para identificar las variables más relevantes para la predicción del desistimiento, que fueron tratadas en el anterior sprint y tienen un buen porcentaje de consistencia y confiabilidad de los datos. En esta etapa se evaluó tres Algoritmos de Segmentación y se ejecutó el PCA de Kmeans de los datos.

Herramientas. Python de Google Colab para los análisis de correlación.

Entregables. Análisis de correlación de variables para la selección de las variables más importantes.

Balanceo del Modelo

Actividades. El equipo realizó el desarrollo del código del modelo predictivo utilizando el lenguaje de programación y las herramientas adecuadas. Se evaluó 11 Modelos analíticos, profundizando sobre Árbol de decisión.

Herramientas. Python de Google Colab para la construcción del código del modelo.

Entregable. Especificaciones del diseño del modelo.

Ejecución del Modelo

Actividades. Se ejecutó los 11 modelos analíticos obteniendo los indicadores más relevantes, para el Árbol de decisión se realizó el ForestCast.

Herramientas: Python de Google Colab para la ejecución del modelo

Entregables. Ejecución de 11 modelos, los cuales sus indicadores son comparados para escoger el mejor método de pronóstico del desistimiento. ForesCast del modelo analítico de Árbol de decisión.

Ajuste del Modelo

Actividades. Se Realizó ajustes y optimizaciones del modelo con el objetivo de mejorar su precisión y rendimiento.

Herramientas. Python de Google Colab para la ejecución del modelo

Entregable: Modelo optimizado.

Evaluación y Validación

Actividades. Se ejecutó pruebas del modelo mejorado y compararlo con otros enfoques de referencia.

Herramientas. Python de Google Colab para la evaluación y validación del modelo.

Entregables. Modelo Evaluado.

Despliegue del Modelo

Actividades. Revisión y validación de la base de datos completa que incluye todas las transacciones de ventas realizadas durante el primer trimestre del año 2024 para Identificar y categorizar con el modelo escogido de árboles de decisiones el desistimiento en tres niveles (Bajo, Medio y Alto).

Herramientas. Python de Google Colab para el despliegue del modelo.

Entregables. Excel que represente visualmente los niveles de desistimiento, para facilitar la interpretación y toma de decisiones por parte de la constructora.

Sprint 5 – Estrategias de Gobernabilidad y Administración del Riesgo (abril 2024 - mayo 2024)

Durante el quinto sprint, el equipo se centró en definir las estrategias de gobernabilidad de los datos para la constructora.

Estrategias de gobierno de datos

Actividades. Se proporcionó a la constructora a partir del nivel de madurez de la gestión de los datos, la ruta estratégica de gobernabilidad para futuros modelos analíticos y eficiencia en los datos de la constructora.

Entregables. Estrategias de gobernabilidad de datos dividido en Dominios de Información.

Administración del Riesgo

Actividades: En esta etapa se evaluó los riesgos identificados en términos de probabilidad e impacto. Priorizando los riesgos y desarrollar estrategias de mitigación.

Herramientas. Matriz de probabilidad e impacto, análisis cualitativo y cuantitativo de riesgos.

Entregables. Matriz de riesgos y plan de mitigación.

Conclusiones y Recomendaciones.

Actividades: Elaboración de conclusiones y recomendaciones basadas en los logros del proyecto, enfocados en la continuidad de la gestión de riesgos, la actualización del modelo, la integración del feedback de los stakeholders, la capacitación del personal y la expansión del modelo a otras áreas.

Entregables. Documento de conclusiones y recomendaciones.

5. Descripción de las Fuentes de Información

La principal fuente de información para este estudio fueron los datos de clientes recopilados y proporcionados por la constructora. Estos datos incluyeron información detallada sobre los clientes que habían desistido desde 2013 hasta la fecha. La base de datos contenía registros de clientes, incluyendo su historial de compras, razones de desistimiento, fechas de desistimiento, entre otros. Estos datos fueron revisados, transformados y analizados para obtener insights relevantes sobre los clientes que habían desistido.

Además de la base de datos de clientes, se utilizaron informes y documentos internos de la constructora. Estos documentos podían incluir informes de ventas, análisis de mercado, estrategias de retención de clientes, entre otros. Estos recursos proporcionaron una visión más amplia de las acciones y decisiones tomadas por la empresa en relación con los clientes que habían desistido, lo que ayudó a contextualizar y comprender mejor los datos recopilados.

Además de las fuentes de información internas, se consultaron una variedad de informes y estudios externos relacionados con el sector de la construcción y la retención de clientes para el

entendimiento del problema y alcance del proyecto. Estos informes provienen de estudios o artículos académicos. Estos recursos proporcionaron una perspectiva adicional para el análisis de los clientes que habían desistido en la constructora

Es importante mencionar que todas las fuentes de información fueron evaluadas según su fiabilidad, relevancia y actualidad. Además, se aplicaron técnicas de análisis de datos adecuadas para extraer conclusiones sólidas y respaldadas por la evidencia obtenida de estas fuentes.

6. Entendimiento del Negocio

En el año 2022, el sector de la construcción en Colombia contribuyó con el 5.3% del PIB del país, convirtiéndose así en uno de los sectores más importantes y representativos. Desde la generación de empleo, estimada en más de 2 millones de trabajadores, hasta la atracción de 2000 millones de dólares de inversión internacional en Colombia (Guerrero, 2023), la industria inmobiliaria ha sido fundamental para el desarrollo económico.

Sin embargo, durante el año 2023, el sector de la construcción se enfrentó a diversos desafíos, entre ellos las altas tasas de interés y las decisiones gubernamentales que redujeron los subsidios para viviendas de interés social. Ante este entorno desafiante, todas las empresas constructoras implementaron nuevas estrategias para mantener sus niveles de venta y rentabilidad.

Por ejemplo, algunas constructoras optaron por estrategias centradas en el cliente y en la experiencia, buscando asegurar un servicio excepcional desde el primer contacto hasta la entrega del proyecto. La transparencia en la comunicación sobre los desafíos financieros y los posibles beneficios a largo plazo de la inversión en propiedades se identificó como un factor clave para generar confianza y fomentar relaciones sólidas con los clientes.

Además, la adaptación a las nuevas tecnologías y tendencias del mercado inmobiliario se convirtió en una prioridad para muchas empresas del sector. La implementación de herramientas digitales para mostrar proyectos, ofrecer recorridos virtuales y facilitar la participación del cliente en el proceso de diseño se identificó como una forma efectiva de mejorar la conexión emocional y la percepción de valor de los proyectos inmobiliarios.

Otro enfoque estratégico importante fue la flexibilidad financiera. Ante las altas tasas de interés, las empresas exploraron opciones innovadoras, como esquemas de financiamiento atractivos, programas de fidelización y descuentos especiales, para hacer más accesible la compra de una propiedad y atraer a potenciales compradores.

En medio de estos desafíos, fue crucial para las empresas constructoras asegurar sus ventas y reducir los desistimientos de clientes. En este sentido, la constructora del presente proyecto, con más de tres décadas de trayectoria, se ha destacado como una figura prominente en el panorama de la construcción y comercialización de viviendas en Colombia. Desde su fundación en 1991, la empresa ha liderado proyectos inmobiliarios significativos en diversas regiones del país, desde la región Andina hasta la región Caribe y el occidente colombiano.

En los últimos años, la constructora ha experimentado un crecimiento continuo y una expansión geográfica significativa. Según su plan estratégico para el año 2020, la empresa acumuló 2,801 unidades en separaciones, lo que representó una inversión de \$481,113 millones de pesos. Estas unidades se distribuyeron en diversas categorías, incluyendo Viviendas de Interés Prioritario (VIP), Viviendas de Interés Social (VIS) y proyectos NO VIS.

La Regional Bogotá, que incluye la capital y municipios cercanos, contribuyó de manera significativa con 1,229 unidades en separaciones y 553 unidades entregadas, consolidando así la presencia y el impacto de la constructora en esta importante área.

Además de su éxito en el ámbito nacional, la constructora ha llevado su compromiso con la calidad y la satisfacción del cliente a nivel internacional con proyectos de construcción en Miami, Florida.

En términos de gobernanza de datos y tecnología, la constructora ha avanzado con la implementación de herramientas digitales, como el Customer Relationship Management (CRM) de Oracle, que centraliza la documentación del cliente y mejora la eficacia en los cierres de negocios. Sin embargo, aún no cuenta con una estrategia sólida de gestión de datos. La empresa también ha destacado en marketing digital con la implementación de Eloqua y la integración con EMABOT para atención al cliente 24/7.

A pesar de los desafíos enfrentados en estos años, especialmente los desistimientos causados por la pandemia del COVID-19, la constructora ha demostrado un enfoque resiliente y un compromiso con la innovación que le ha permitido adaptarse y seguir construyendo sueños para las familias colombianas, como lo establece su misión. Para mejorar la experiencia del cliente, la constructora está enfocada en comprender y atender las necesidades del perfil específico de sus clientes. Esto implica una atención especial a los detalles que preocupan a los compradores de vivienda, desde las características específicas de los inmuebles hasta los aspectos financieros que pueden influir en sus decisiones. La empresa reconoce la importancia de abordar la problemática de los desistimientos en los negocios y está trabajando activamente en estrategias para reducir esta incidencia; teniendo en cuenta esta premisa, se realiza un diagnóstico inicial.

Como se observa en la Figura 2, la constructora contaba con una tasa de desistimientos del 30.34% durante los últimos 10 años, con una meta de 20% para el desistimiento, se observó

que las causas más recurrentes son la incapacidad económica del comprador, la pérdida de empleo y negación de créditos y subsidios, y la Cartera en mora.

Figura 2

Tablero del diagnóstico inicial



Nota: Indicadores iniciales del estado actual de la constructora.

En los últimos años, se ha notado una mejora considerable en la gráfica comparativa de desistimientos frente a las ventas en la constructora. Este progreso se atribuye a la implementación de términos contractuales y promesas de compraventa favorables, A pesar de estos logros internos, no se puede pasar por alto la notable influencia que aún ejerce el entorno externo que presenta los posibles compradores.

Es evidente que una proporción significativa de clientes potenciales que decidió abandonar sus compras lo hace en respuesta a factores macroeconómicos y a las condiciones generales del sector de la construcción. En este contexto, se decidió acotar y filtrar las bases de datos a los últimos cinco años. Esta estrategia permitió no solo examinar detenidamente el

historial de desistimientos y ventas, sino también entender de manera más profunda aspectos financieros en el comportamiento de los clientes.

La unificación de datos de ventas, recaudo, perfilamiento de clientes y desistimientos constituye un avance significativo hacia el análisis integral de las operaciones de la constructora. En este proceso de consolidación de información, el objetivo es identificar y resaltar las variables más significativas que posibiliten un modelado de datos más preciso.

El gráfico reveló una clara necesidad de abordar la predicción temprana del desistimiento del comprador por factores internos y externos que afectan a la hora de tomar una decisión de compra. Por lo anterior, el presente proyecto nace de la necesidad de la constructora para crear estrategias útiles de retención de acuerdo con una adecuada gestión de datos, el cual estableció bases para un análisis más profundo y preciso, crucial para la toma de decisiones y la mejora continua.

7. Descripción de la Problemática y Propuesta Resolutiva

El desistimiento en proyectos de vivienda es un problema crítico para las constructoras, particularmente en el contexto colombiano. Este fenómeno, genera pérdidas económicas significativas, afecta la planificación de proyectos y reduce la confianza en el mercado inmobiliario. Según datos del Departamento Administrativo Nacional de Estadística (DANE), el sector de la construcción ha experimentado un aumento en las tasas de desistimiento en los últimos años, afectando tanto a empresas como a compradores.

Análisis Detallado del Problema

El desistimiento puede ser causado por una variedad de factores, incluidos problemas financieros de los compradores, cambios en las condiciones del mercado, falta de confianza en la constructora, o insatisfacción con el inmueble. En el contexto colombiano, se observa que

aproximadamente el 15% de los contratos preliminares de compra de vivienda terminan en desistimiento, lo que representa una pérdida significativa para las constructoras tanto en términos económicos como de reputación.

Causas y Factores Contribuyentes

Entre las principales causas identificadas se encuentran:

Problemas Financieros, Pérdida de empleo o cambios en la situación financiera de los compradores.

Condiciones del Mercado, Variaciones en las tasas de interés hipotecario y fluctuaciones en los precios de los inmuebles.

Confianza y Satisfacción, Percepción de la calidad del inmueble y la reputación de la constructora.

Evidencia del Problema

Un estudio realizado por la Cámara Colombiana de construcción (Camacol) en 2022 reveló que las constructoras reportaron pérdidas de hasta el 10% de sus ingresos proyectados debido a desistimientos. Estos datos subrayan la necesidad de contar con herramientas que permitan anticipar y mitigar este riesgo.

Propuesta Resolutiva

Se propone el desarrollo de un modelo predictivo de desistimiento, basado en técnicas de aprendizaje automático, que permita a las constructoras identificar a los clientes con mayor probabilidad de desistir un año después de la separación del inmueble. El modelo utilizará datos históricos y en tiempo real sobre las características demográficas, financieras y comportamentales de los compradores, así como factores macroeconómicos y específicos del proyecto de vivienda.

La implementación de un modelo predictivo para prevenir desistimientos en el sector de construcción requiere un enfoque integral que abarque diversas variables y factores. Aquí se detallan los pasos y estrategias que el modelo utiliza:

Recopilación y Análisis de Datos

Recopila datos relevantes de clientes que hayan desistido de compras anteriores y aquellos que no. Se deben analizar datos financieros, comportamientos de crédito, historiales de transacciones y cualquier otro indicador relevante.

Identificación de Patrones y Variables Significativas

Utilizando técnicas de análisis de datos, el modelo debe identificar patrones y variables significativas que puedan estar asociadas con los desistimientos. Esto incluye variables como Salario, Edad, Cargo, Profesión, y otros factores relevantes. El modelo también debe considerar la interacción entre variables para obtener una comprensión más completa.

Segmentación del Perfil del Cliente

Basándose en los resultados del análisis, el modelo segmenta el perfil del cliente en grupos que compartan características similares. Esto permitirá personalizar las estrategias de retención para abordar las necesidades específicas de cada cliente.

Evaluación Continua y Actualización del Modelo

El modelo predictivo no es estático y debe someterse a evaluación y actualización continua. Se deben incorporar nuevos datos a medida que estén disponibles y ajustar el modelo según las tendencias cambiantes del mercado y las preferencias del cliente. La retroalimentación de clientes que han desistido puede ser muy valiosa para mejorar la efectividad del modelo con el tiempo.

La efectividad del modelo será evaluada mediante la tasa de aciertos en la predicción de desistimientos y la reducción real de los mismos tras la implementación de medidas preventivas basadas en las predicciones. Indicadores clave incluirán la disminución en la tasa de desistimiento y el aumento en la satisfacción y retención de clientes.

Posibles Obstáculos

Se anticipan desafíos como la resistencia al cambio organizacional y la necesidad de inversiones iniciales en tecnología y capacitación. Además, la precisión del modelo puede verse afectada por la calidad y disponibilidad de los datos.

Implicaciones y Repercusiones del Modelo

La implementación exitosa del modelo predictivo no solo reducirá las tasas de desistimiento, sino que también mejorará la confianza de los compradores y fortalecerá la posición competitiva de la constructora en el mercado. A largo plazo, esto podría contribuir a una mayor estabilidad en el sector inmobiliario colombiano. Comparado con métodos tradicionales basados en intuición y experiencia, el uso de un modelo predictivo ofrece una mayor precisión y permite una gestión más eficiente y proactiva de los riesgos de desistimiento.

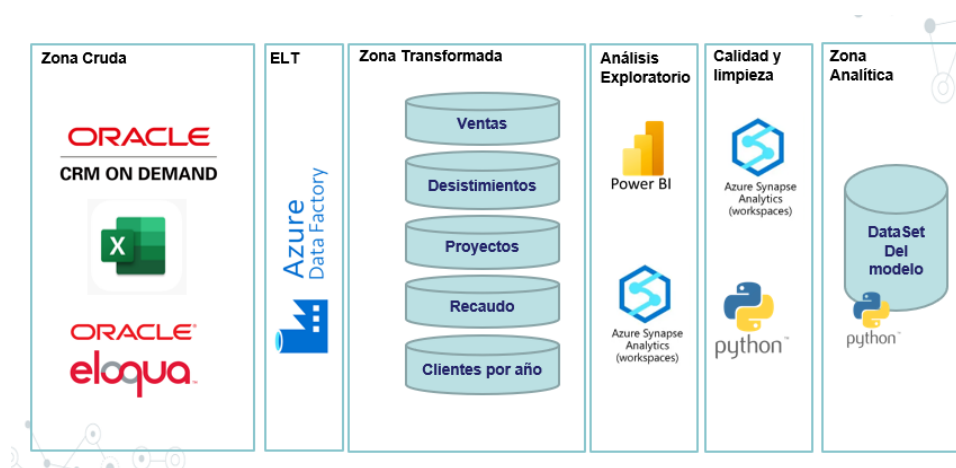
La problemática del desistimiento en proyectos de vivienda presenta un desafío significativo para las constructoras colombianas. La propuesta de un modelo predictivo de desistimiento representa una solución innovadora y viable, con el potencial de transformar la gestión de riesgos en el sector. Al anticipar y mitigar los desistimientos, las constructoras podrán mejorar su eficiencia operativa y fortalecer la confianza de los compradores, contribuyendo así al desarrollo sostenible del mercado inmobiliario.

8. Arquitectura Del Modelo De Datos

En la arquitectura del modelo de datos para el ejercicio analítico se realizó la identificación del origen de los datos a utilizar, detallando desde su captura de entrada hasta la disponibilidad en el dataset. Como se observa en la Figura 3.

Figura 3

Arquitectura del modelo analítico propuesto



El flujo se divide en varias zonas y etapas, cada una con un propósito específico dentro del ciclo de vida de los datos.

En la Zona Cruda se recopiló toda la información utilizada en el proyecto, estos datos provienen de sistemas de información robustos como el CRM o sencillas como Excel, compartidas por las áreas de negocio. En esta etapa de captura de datos se desarrolló y ejecutó un ETL con un pipeline en Data Factory de Azure para almacenar la información y las transformaciones pertinentes.

En la Zona Transformada, se ejecutó el diagnóstico de calidad, la limpieza, análisis exploratorio, perfilamiento de los datos divididos en dominios de información de ventas, desistimientos, proyectos, recaudos y clientes.

En el Análisis Exploratorio de los datos con Power BI y Synapse de Azure se realizó las pertinentes validaciones iniciales con las que se construyeron el perfil del desistido y comprador, base para realizar la primera selección de variables importantes para el modelo.

En la Zona Analítica, se ejecutó el dataset con las variables priorizadas en el proceso de transformación y las cuales cuenta con una calidad pertinente.

9. Análisis Exploratorio

Al analizar los datos ya transformados y cargados en Azure, se realizó un diccionario de datos para recoger los metadatos con las descripciones específicas de las variables, actuando como referencia para entender los datos y comprensión precisa de la integridad referencial de los dominios de información, se revisaron los campos participantes en las fuentes de información priorizando las más representativas para el modelo analítico.

En la tabla 2 se detallan las variables más relevantes de la base de datos, las cuales fueron evaluadas para posteriormente ejecutar el proceso de calidad. Los campos incluidos contienen información crucial, como el nombre del proyecto macro y el nombre específico de la manzana, que permiten identificar las diferentes áreas dentro de los macroproyectos.

Tabla 2

Diccionario de datos

Orden	Nombre Variable.	Nombre en base de datos	Detalle variable.	Captura
1	Macroproyecto	Macroproyecto	Nombre del proyecto Macro	Selección
2	Proyecto	Proyecto	Nombre de la manzana	selección
3	agrupación	agrupación	Numero de torre y apartamento	viene de otro proceso
4	Cod	Cod	Código interno	viene de otro proceso
5	Tipo	Tipo	Tamaño y tipo de vivienda	viene de otro proceso
6	Fecha Rad Vta	Fecha Rad Vta	Fecha de la separación	Digitado

7	Casa/Apt	Casa/Apt	Numero de torre y apartamento	viene de otro proceso
8	Garaje	Garaje	Si tiene parqueadero propio	viene de otro proceso
9	Deposito	Deposito	Si tiene deposito propio	viene de otro proceso
10	Reformas Si/No	Reformas Si/No	Si el inmueble es atípico	viene de otro proceso
11	Comprador	Comprador	Nombre del comprador	Digitado
12	Cedula	Cedula	identificación del comprador	Digitado
13	Tipo Venta	Tipo Venta	tipo de pago del inmueble	selección
17	Entidad crédito	Entidad crédito	Entidad con la que se solicita el restante del pago del inmueble	selección
18	Entidad Subsidio	Entidad Subsidio	Entidad con que se tramita subsidio	selección
31	Vr_Neto	Vr_Neto	Valor del inmueble	viene de otro proceso
32	Bono	Bono	Bono de descuento por escrituración	Digitado
33	crédito	crédito	crédito hipotecario de la vivienda	aplicación calculada
34	C_Ter	C_Ter	crédito hipotecario de la vivienda con una entidad diferente al crédito constructor	aplicación calculada
35	C_I	C_I	Créditos terceros	aplicación calculada
36	Subsidio	Subsidio	Valor del subsidio a desembolsar	aplicación calculada
38	Ahorro Programado	Ahorro Programado	Valor del ahorro programado del cliente	Digitado

El diccionario de datos proporciona una visión detallada de los campos y variables (Dong-Hui Jin, 2018) utilizados en la base de datos de la constructora, particularmente relevantes para el análisis de desistimientos, como el tipo y tamaño de vivienda, la fecha de separación, el estado de la vivienda en términos de reformas o características atípicas, así como detalles sobre los compradores y sus formas de financiamiento.

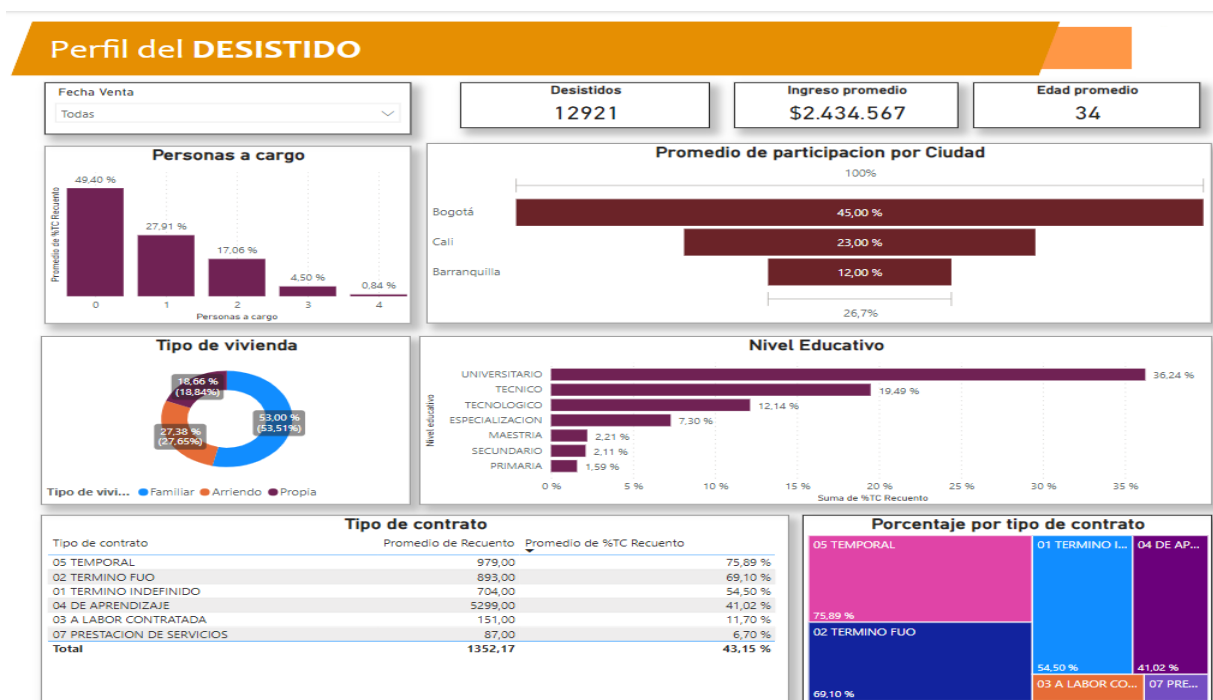
Se organizó y analizó la información para identificar patrones que demostraron cómo la calidad de las bases de datos puede influir en el ingreso manual de datos por parte de los agentes

en las salas de ventas. Este proceso generaba inconsistencias y errores en los datos, lo cual impedía que las tablas reflejaran de manera completamente precisa la realidad. Para abordar esta situación, se planteó a la constructora diferentes estrategias para limitar la variabilidad en los formularios de recopilación de datos de los clientes, mediante la estandarización y homogeneización del proceso de captura de datos, lo cual ayudará a mejorar la calidad y la confiabilidad de la información. Estas medidas se orientaban a minimizar los errores, garantizar una mayor consistencia en los registros y facilitar un análisis más preciso de los desistimientos en la constructora, especificadas en el capítulo de recomendaciones de este documento.

Perfil del Desertor, con la información recopilada sobre el perfil del cliente que decide abandonar el negocio con la constructora, se exploró en mayor detalle en el siguiente tablero explicado en la Figura 4.

Figura 4

Perfil del Desertor



Nota: Información de variables importantes para el entendimiento del desertor.

La exploración detallada de los resultados del análisis reveló que el 30,34% de clientes que decidió abandonar un proyecto con la constructora suele tener entre 31 y 40 años, con una estabilidad financiera de ingresos mensuales promedio de \$2.434.567 COP. Este grupo esta mayormente compuesto por personas solteras, sin hijos, su nivel educativo universitario y la vinculación laboral a través de contratos de aprendizaje o temporales. Estos hallazgos dan una visión más clara del perfil del desertor que se alejan de los servicios de la constructora, sirviendo de base para estrategias que buscan retenerlos de manera más efectiva en el futuro.

Perfil del Cliente, La exploración de la base de datos reveló características distintivas del perfil del comprador que opta por adquirir una vivienda a través de la compañía, ver Figura 5.

Figura 5

Perfil del comprador



Nota: Información de las variables importantes para el comportamiento del comprador.

Los compradores presentan un ingreso promedio de 5.477.437. Predominando con un 30% los clientes pertenecientes al grupo demográfico de 31 a 50 años. Este segmento se caracteriza por no tener hijos o con un único hijo. Adicional, se observó porcentajes representativos del 27% y 22% de compradores con contratos a término fijo o indefinido.

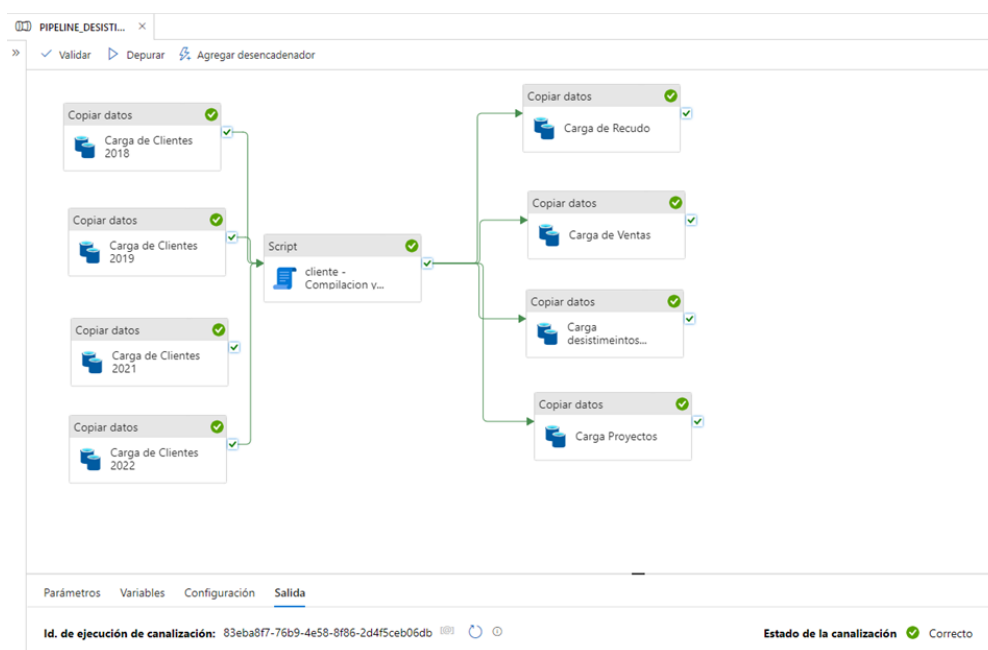
El estado civil de los clientes muestra una diversidad considerable, abarcando desde solteros, quienes representan un segmento importante del 50%, hasta clientes casados. En términos de vivienda, los clientes tienden a residir en hogares familiares o a optar por viviendas arrendadas, destacando las preferencias habitacionales que influyen en sus decisiones de compra.

Este análisis detallado proporcionó información de las preferencias habitacionales de los compradores y su perfil socioeconómico. Estos hallazgos son esenciales para el desarrollo de estrategias futuras y mejoran la comprensión de cómo la constructora puede adaptarse mejor a las necesidades y expectativas cambiantes de su base de clientes.

10. Preparación De Datos

Construcción Del Pipeline

Para la adecuada explotación y análisis de los datos se creó una fuente unificada dimensional por Data Factory, con el objetivo de realizar el proceso de calidad de las fuentes y creación del dataset del modelo, ver figura 6.

Figura 6*Pipeline Datafactory*

Nota: Flujo del proceso de la ETL en DataFactory

El pipeline implementado en Azure Data Factory se encarga de la carga de la información cruda del modelo, estructurándose en las siguientes etapas:

Carga Clientes, En esta fase, se realiza la carga de las bases de datos de clientes, seleccionando específicamente los años que serán evaluados en el modelo. Se excluyó el año 2020 debido a su naturaleza atípica. La información recopilada se consolidó de manera eficiente en una única tabla de datos de clientes, proporcionando así una base integral para el análisis.

Carga Recaudo, Este proceso incorpora la información contable relacionada con la cantidad recaudada por cliente y proyecto en su totalidad. La inclusión de datos financieros es esencial para la comprensión completa de la situación económica de los proyectos y clientes asociados.

Carga Ventas, Se cargó la información detallada sobre el proceso de venta, segmentada por cada proyecto. Este paso permite analizar en profundidad las dinámicas de venta y contribuye a la comprensión de los factores que impactan en el rendimiento de cada proyecto.

Carga Desistimientos, En esta etapa, se incorporó información histórica relacionada con desistimientos de proyectos en la constructora. Este análisis retrospectivo proporciona insights valiosos para comprender las razones detrás de los desistimientos y permite ajustar estrategias para minimizar dichos eventos en proyectos futuros.

Carga Proyectos, Por último, se llevó a cabo la carga de información básica de los proyectos en venta. Esta fase se centró en recopilar datos esenciales para una visión general de los proyectos, sirvió de base para la evaluación global del rendimiento y la toma de decisiones estratégicas.

La implementación de este pipeline aseguró la integración sistemática y completa de datos crudos, preparando el terreno para cualquier análisis predictivo efectivo en el ámbito de la evaluación de proyectos de la constructora. Como conclusión a esta primera transformación y evaluación de los datos de la compañía, se obtuvo un diagnóstico inicial, que nos permitió visualizar de manera global como se encuentra la constructora en sus sistemas de retención de clientes.

Perfilamiento de Datos

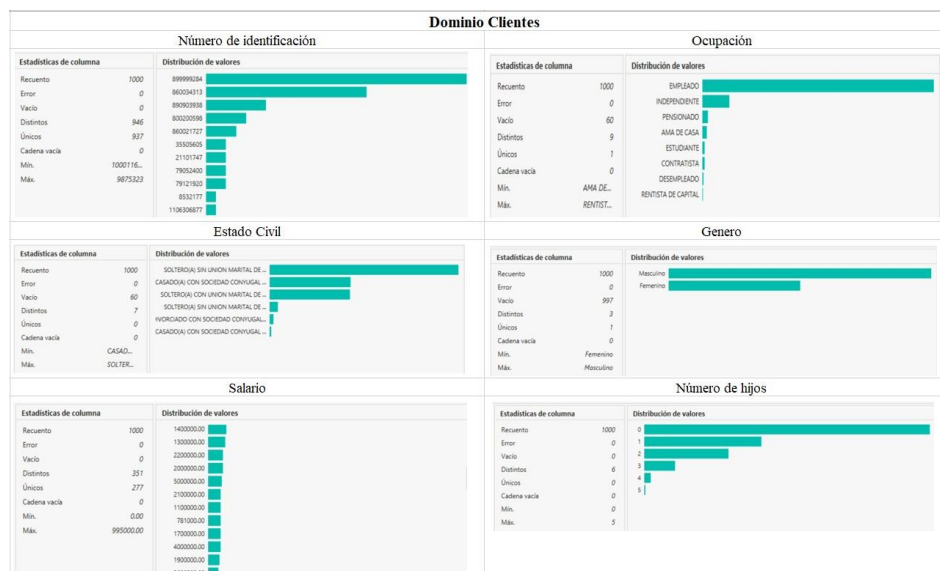
Seleccionadas las variables más relevantes del modelo se hizo un perfilamiento estadístico inicial con el objetivo de ejecutar limpieza de los datos (Ulrike, 2015), este perfilamiento se hizo sobre una muestra de 1.000 registros en Power Bi, arrojando la siguiente información estadística.

Dominio Clientes

En este dominio se realizó el perfilamiento a las variables más relevantes de los datos del cliente, ver figura 7.

Figura 7

Perfilamiento del dominio clientes



Nota: Información estadística de los datos recolectados.

Número de Identificación Identificador único para cada cliente, La integridad de este campo es crucial para el seguimiento efectivo de cada cliente en la base de datos de la constructora.

Ocupación, Ofreció información importante sobre la profesión del cliente, Este dato facilita la personalización de productos y servicios según las necesidades y características específicas de cada grupo.

Salario, Indicó el nivel de ingresos del cliente, Se examinaron las estadísticas descriptivas para entender la distribución, el mínimo y el máximo de los ingresos. Este análisis ayuda a adaptar estrategias de precios y promociones.

Género, Facilitó la segmentación y personalización de servicios según el género del cliente, Se evaluó los datos de género para identificar patrones.

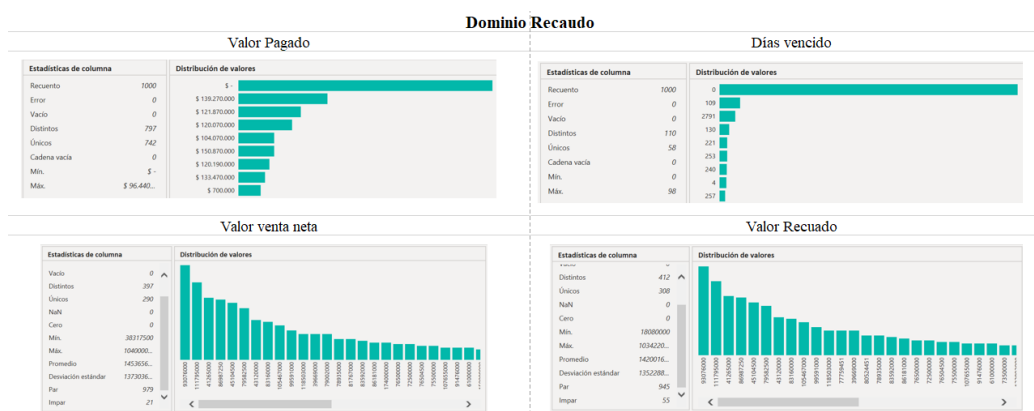
Estado Civil Se analizó la distribución de estados civiles para identificar oportunidades de personalización en productos y servicios específicos para cada situación familiar.

Dominio Recaudo

En este dominio se realizó el perfilamiento a las variables más relevantes de los datos del recaudo del proyecto vendido, ver figura 8.

Figura 8

Perfilamiento de dominio de recaudo



Nota: Información estadística de los campos de recaudo

Valor de Venta, Indicó el monto total de las ventas después de deducciones, Este gráfico presentó una amplia dispersión de valores, con 357 distintos registrados y una concentración más pronunciada en valores inferiores, lo que refleja una tendencia a menores montos de transacciones para la mayoría de las ventas

Valor Recaudado, Representó la cantidad de dinero recogido, se evaluó la relación entre el valor recaudado y el valor de venta neta para medir la capacidad de pago del comprador de la

vivienda. El valor recaudado mostró una tendencia descendente en el total recaudado según se alejaba de los valores más altos.

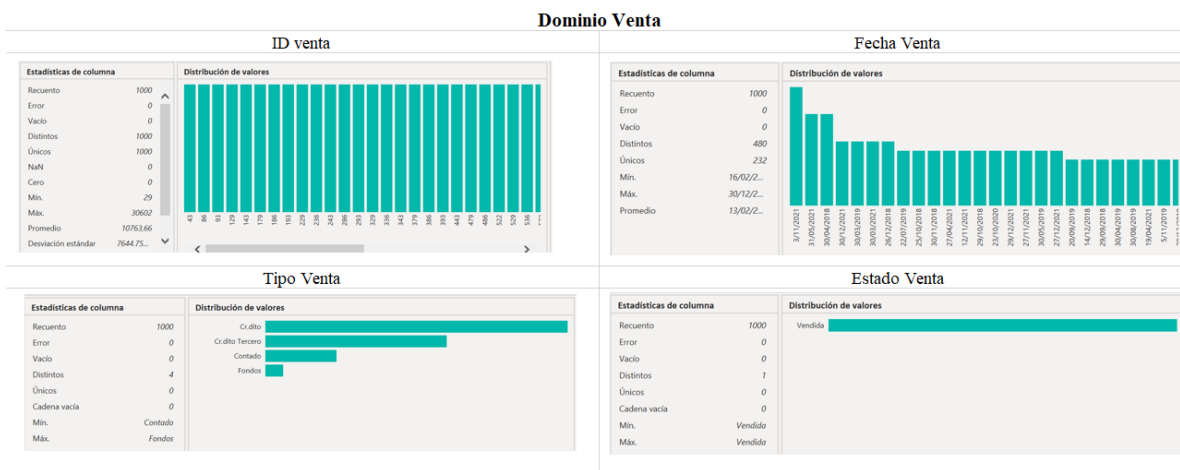
Días de Mora, esta variable indicó la cantidad de días que un pago ha estado vencido, mostraron una distribución donde la mayoría de los valores estaban concentrados en intervalos inferiores, con un máximo de 255 días.

Dominio Ventas

En este dominio se realizó el perfilamiento estadístico a las variables más relevantes de las ventas de la constructora. ver figura 9.

Figura 9

Perfilamiento de dominio de ventas



Nota: Información estadística de los campos de ventas

Fecha Venta, Indica la fecha en la que se realizó la venta, Se realiza un análisis temporal que incluye medidas como la media, la mediana y la desviación estándar de las fechas de venta. Esto permite identificar tendencias estacionales, períodos de mayor actividad y posibles patrones temporales en las transacciones.

Tipo de Venta, Esta variable indicó la categoría o modalidad de la transacción, Se realizó un análisis de frecuencia para entender la distribución de los tipos de venta. Esto proporciona información sobre las preferencias del cliente.

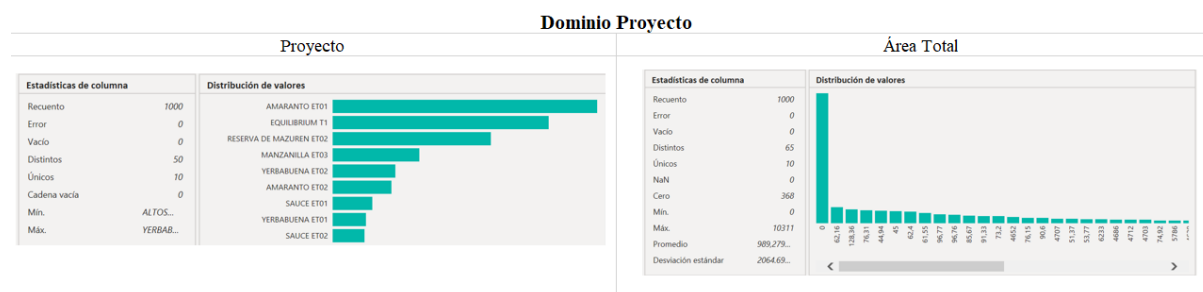
Estado Venta Indica el estado actual de la transacción, Se evalúa la proporción de diferentes estados de venta.

Dominio Proyecto

En esta etapa se realizó el perfilamiento estadístico a las variables más relevantes de los proyectos de la constructora. ver figura 10.

Figura 10

Perfilamiento dominio de proyecto



Área Total, Representa la dimensión total del terreno o construcción en metros cuadrados, se examina la distribución de áreas totales para entender la diversidad de tamaños entre los diferentes proyectos.

Limpieza De Datos

A continuación, se detallan las acciones de limpieza de datos específicas realizadas para cada variable en el conjunto de datos:

Cargo, se ejecutó verificación de consistencia y corrección de posibles errores tipográficos. Se corrigió codificación adecuada de categorías para su inclusión en el modelo.

Número Identificación Comprador, Se validó la unicidad y consistencia en los números de identificación y manejo de duplicados y registros inconsistentes.

Estado Civil, se realizó Identificación y manejo de categorías no válidas o desconocidas y codificación adecuada para la inclusión en el modelo.

Fecha Nacimiento, Cálculo de la edad a partir de la fecha de nacimiento.

Validación y tratamiento de valores atípicos o inconsistentes.

Ingreso Promedio Mensuales, Imputación de valores faltantes con técnicas de la media y normalización para garantizar comparabilidad entre diferentes escalas salariales.

Nivel Académico, Manejo de categorías no válidas o desconocidas y codificación adecuada para su inclusión en el modelo.

Número de Hijos, Tratamiento de valores nulos y manejo de posibles inconsistencias.

Salario, Imputación de valores faltantes y tratamiento de outlier y normalización garantizando la comparabilidad entre diferentes niveles salariales.

Género, Identificación, manejo de categorías no válidas o desconocidas y codificación adecuada para su inclusión en el modelo.

Días de Mora, Tratamiento de valores negativos o inconsistentes y normalización para garantizar comparabilidad.

Valor Recaudo, Manejo de valores nulos y tratamiento de posibles outliers.

Valor Venta Neta, Validación y tratamiento de valores inconsistentes.

Calidad De Datos

Tras llevar a cabo la limpieza de datos en las bases de datos y seleccionar las variables clave para el modelo, se realizó la creación de un dashboard de data quality scorecard que incorpora métricas de consistencia, validez y unicidad, como se puede observar en la Figura 11.

El análisis realizado reveló que tras este proceso de perfilamiento y limpieza se lograron indicadores de calidad superiores del 90 % para 5 de los 6 dominios de información. Este logro no solo refleja un alto nivel de integridad y precisión en los datos, sino que también contribuye significativamente a la efectividad del modelo analítico.

Figura 11

Data Quality Scorecard



El tablero de control de calidad, ver figura 11, proporciona una visión holística de la fiabilidad y validez de los datos utilizados en el modelo. La métrica de consistencia evaluó la uniformidad y coherencia de los datos, asegurando que no haya discrepancias significativas. La validez se centró en la precisión y relevancia de la información, garantizando que los datos reflejen con precisión la realidad que representan. La unicidad abordó la presencia de duplicados, asegurando que cada dato sea único y no redundante.

Al alcanzar un nivel de calidad superior al 90%, se fortalece la confianza en la solidez del modelo analítico. Esto se traduce en resultados más confiables y precisos, lo que mejora la capacidad de la organización para tomar decisiones informadas basadas en datos de alta calidad.

La combinación del proceso de perfilamiento y la limpieza efectiva de los datos ha elevado la calidad general de la información utilizada. Este logro no solo permitió optimizar el rendimiento del modelo, sino que también ayudará a la constructora para aprovechar al máximo las capacidades predictivas y descriptivas del análisis de datos en el futuro.

Ventanas de Tiempo LAG

En la preparación de los datos para el modelo, se crearon los “Lags” que son variables que tiene como objetivo capturar y analizar la relación atemporal entre las diferentes características asociadas al desistimiento de un inmueble. Al desplazar los valores en un año, estamos considerando cómo las condiciones y comportamientos pasados pueden influir en las observaciones actuales o futuras.

Se seleccionaron las variables y se creó una muestra de 1000 registros validando la consistencia, duplicidad y validez de los datos ya procesados. Estas variables se describen en la Tabla 3.

Tabla 3

Variables ventanas de tiempo

Dominio de negocio	Variable seleccionada	Descripción de la variable
Proyecto	Área Total	Superficie de lote construido
Cliente	Cargo	Cargo al cual el comprador pertenece
Recaudo	Días de mora	Número de días incumplidos en cuotas del plan de pagos
Ventas	Estado	Estado en que se encuentra la unidad (Vendido, Separado, disponible)

Cliente	Estado Civil	Situación legal del comprador respecto a obligaciones y circunstancias personales
Ventas	Tipo venta	Característica del tipo de venta del inmueble
Ventas	Fecha de Venta	Fecha de la separación del inmueble en sala de ventas
Desistimiento	Fecha Desistimiento	Fecha en que el comprador desiste de la comprar del inmueble
Nueva variable	Porcentaje de recaudo	Porcentaje de recaudo del inmueble separado a un año
Nueva variable	Edad	Edad del comprador principal calculada con la fecha de nacimiento
Recaudo	Pagos Inmueble	Cantidad de pagos realizados
Cliente	Salario	Promedio de Ingreso que recibe el comprador anualmente
Cliente	Genero	Genero del comprador
Recaudo	Valor pagado	Recaudo aportado por el comprador
Recaudo	Valor Venta Neta	Valor de venta del inmueble
Proyecto	Tipo de vivienda	Tipo de vivienda del inmueble
Nueva variable	Desistimiento	Flag de desistimiento por cliente e inmueble

Nota: Variables seleccionadas para realizar las ventas Lag.

Una vez identificada las variables, se procede a cargar la muestra a Colab Python, validando la consistencia del dataset.

Se ejecutó la transformación de la variable Fecha de venta, la cual es la variable seleccionada para crear los Lags. Se definió los periodos de los 4 Lags creados y se crea un dataset con todas las variables creadas. La creación de Lags proporcionó una perspectiva enriquecida sobre la evolución de los datos, permitió una exploración más profunda, perfilada y precisa sobre de las dinámicas asociadas a la deserción de clientes inmobiliarios. Como se observa en la Figura 12.

Figura 12*Ventanas Lags*

	Enero 2018 - Enero 2019	Enero 2019 - Enero 2020	Enero 2021 - Enero 2022	Enero 2022 - Enero 2023
Lag 1	Desistidos Lag 1			
Lag 1		Desistidos Lag 2		
Lag 3			Desistidos Lag 3	
Lag 4				Desistidos Lag 4

Nota: Lags de la información del perfil del desistido de los últimos 4 años.

11. Creación Del Modelo

Para la creación del modelo, se utilizó toda la información descrita, teniendo en cuenta las ventanas de tiempo y las variables más importantes, con el fin de predecir, de manera clara, la probabilidad de desistimiento en baja, media o alta. (Collins, 2019) Siendo alta la categoría con mayor probabilidad que presenta el desistimiento.

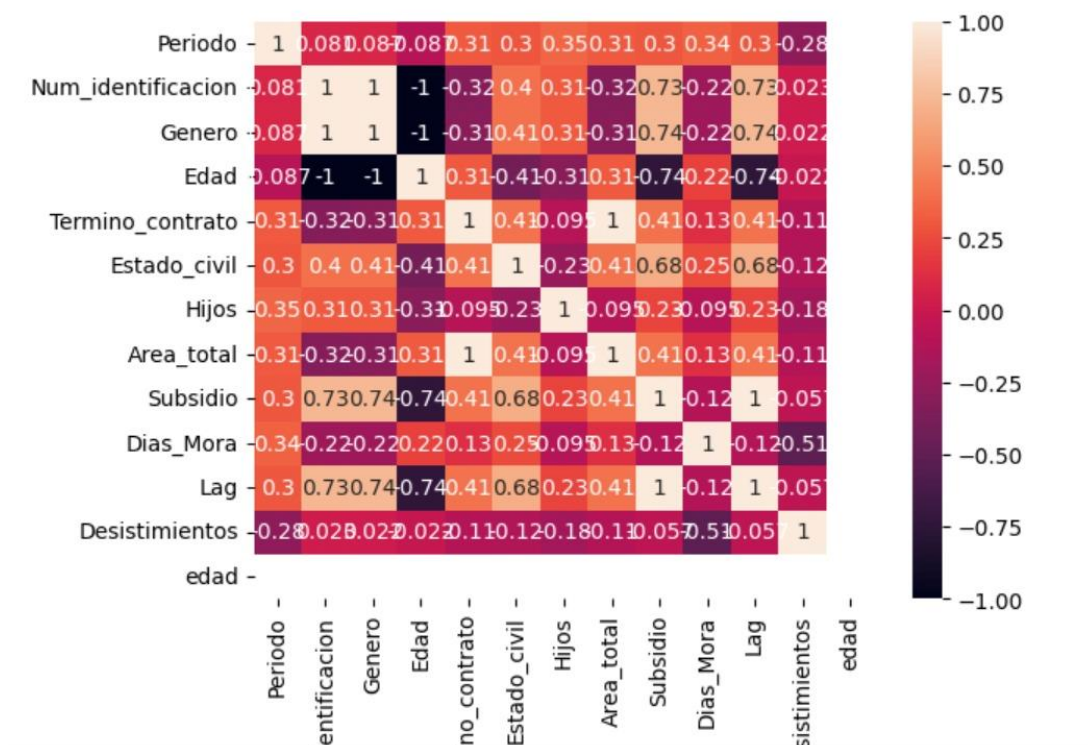
Selección De Variables

El coeficiente de correlación de Pearson, una métrica estadística fundamental, permite evaluar la relación lineal entre dos variables. Su rango de valores va desde -1 hasta 1, donde -1 indica una correlación negativa perfecta, 1 una correlación positiva perfecta y 0 la ausencia de correlación lineal. La interpretación de este coeficiente es crucial en el análisis de datos, ya que proporciona información sobre la dirección y la fuerza de la relación entre las variables consideradas. No es necesario que los datos sigan una distribución lineal para calcular el coeficiente de correlación de Pearson, ya que se basa en las covarianzas y varianzas de las variables en cuestión. Por lo tanto, se utiliza en la etapa descriptiva del análisis para explorar si existe una relación lineal entre las variables de interés, lo que ayuda a comprender mejor la estructura de los datos y a identificar posibles patrones o tendencias.

Para el modelo se analizaron todas las variables descritas en la tabla 3, incluyendo la transformación y limpieza descrita en la sección de calidad de datos. Se realizó una matriz de correlación, esto con el objetivo de seleccionar las variables que más impactan desistimientos, de manera numérica (Zhao, 2023), ver figura 13.

Figura 13

Grafica de correlación de variables



En esta correlación inicial, se observó para la variable “Desistimientos” no se tiene una correlación fuerte o totalmente proporcional con alguna variable, pero tiene relación con la edad, el salario, el número de hijos, la capacidad de pago, el valor del inmueble. Por lo que la variable objetivo fue analizada con las variables más importantes (Ivonne Roy-García, 2019).

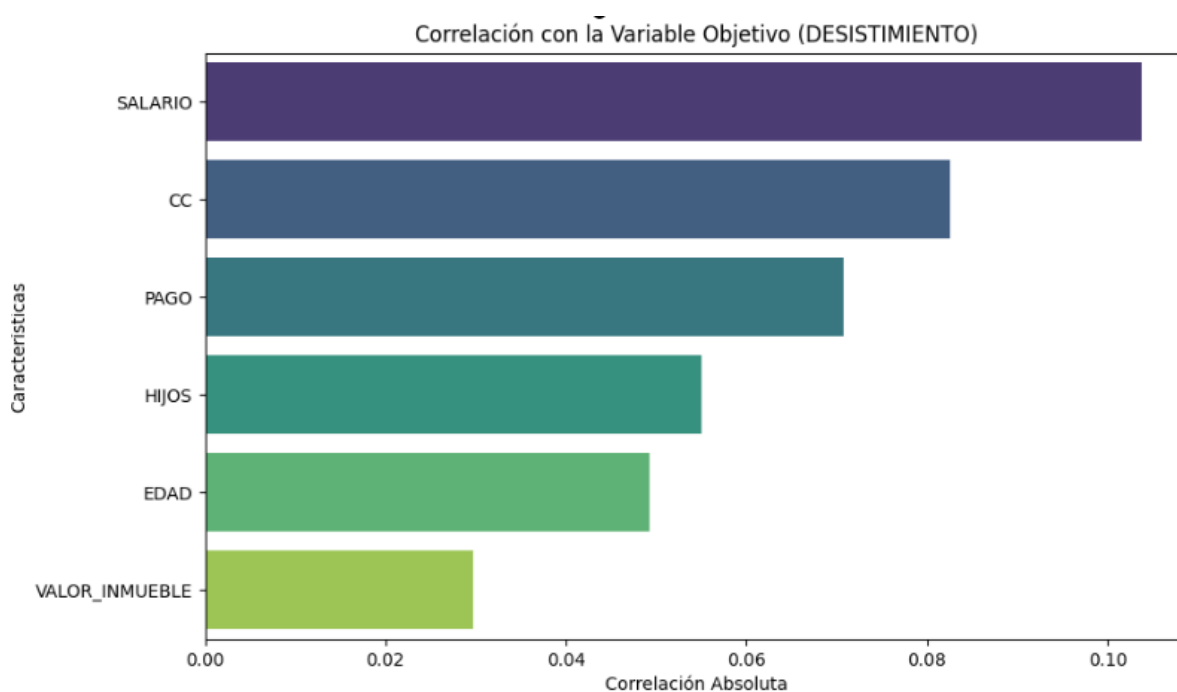
La utilización de correlaciones efectivas asegura la calidad del modelo predictivo al identificar las variables más relevantes y significativas para la predicción del resultado objetivo. Al analizar la fuerza y la dirección de las asociaciones entre las variables, podemos seleccionar aquellas que tienen un impacto significativo en el resultado objetivo y descartar aquellas que tienen una relación débil o insignificante.

En el análisis de la variable objetivo en relación con las variables predictoras, se empleó la correlación de Pearson para evaluar la fuerza y la dirección de la relación lineal entre ellas.

Esta elección se basó en su capacidad para medir la relación entre variables continuas, lo que resultó apropiado dado que algunas de las variables en estudio eran de naturaleza cuantitativa. La correlación de Pearson proporcionó una métrica numérica que permitió identificar la intensidad y la dirección de las asociaciones entre la variable objetivo y las variables predictoras. Al correr el código se observa que las variables finales son las que se observan en la Figura 14.

Figura14

Gráfica de correlación absoluta



Segmentación de Variables

La edad, el salario, el número de hijos, la capacidad de pago y el valor del inmueble fueron las variables seleccionadas para la ejecución del modelo con la variable objetivo de “Desestimiento”. Con esta información, se ejecutó una evaluación de segmentación por clustering, para la ejecución del presente proyecto se seleccionaron tres de los algoritmos más reconocidos con el objetivo de identificar el método más adecuado para el modelo analítico

propuesto. Los algoritmos evaluados fueron KMeans, DBSCAN y MeanShift, aunque actualmente existen numerosas metodologías, estas tres fueron escogidas por su fácil adaptabilidad a cualquier análisis de segmentación.

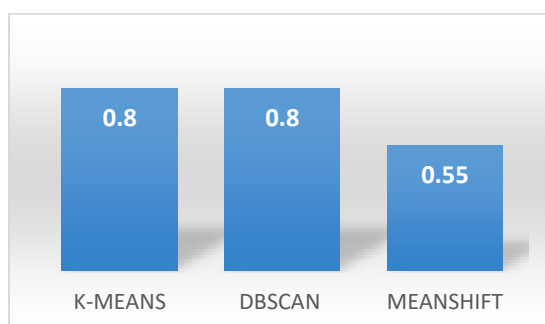
En el contexto de modelos de desistimiento, una buena segmentación ayuda a identificar grupos de clientes con diferentes probabilidades de desistimiento, permitiendo la implementación de estrategias de retención más efectivas.

En el análisis de datos para identificar patrones de desistimiento de clientes, es fundamental seleccionar la técnica de clustering más adecuada. La selección de la técnica se basa en la forma y la dispersión de los datos. El objetivo de la evaluación de estas técnicas en la selección de variables fue la de categorizar a los clientes en grupos homogéneos para identificar estrategias específicas al stakeholder del proyecto, así mismo dar un análisis descriptivo de la tendencia de los datos.

Se llevó a cabo la ejecución de los tres algoritmos de clustering, obteniendo como resultados que las mejores puntuaciones de silueta, indicador que evalúa la calidad de los clústeres, fueron generadas por los algoritmos KMeans y DBSCAN. Esta comparación se visualiza en la Figura 15.

Figura 15

Indicador Silueta de Algoritmos de Clustering



En la Tabla 4 se detalla las ventajas y desventajas de cada algoritmo ejecutado de acuerdo a los resultados, lo que ayudo al equipo a escoger el mejor método para realizar el análisis descriptivo de las variables seleccionadas para el modelo.

Tabla 4

Comparación de Algoritmos de Clustering

Punto para evaluar	KMeans	DBSCAN	MeanShift
<i>Puntuación de Silueta</i>	0.80	0.80	0.55
<i>Ventajas</i>	Fácil de interpretar, Indica que los clústeres están bien definidos y separados. La mayoría de los puntos están correctamente agrupados	Identificó cuatro clústeres con buenos resultado	Identificó dos clústeres principales, lo que sugiere una agrupación insuficiente.
<i>Desventajas</i>		Algunos puntos no fueron agrupados	Menos efectiva en segmentar los clústeres presentes en los datos. Los clústeres no están bien definidos

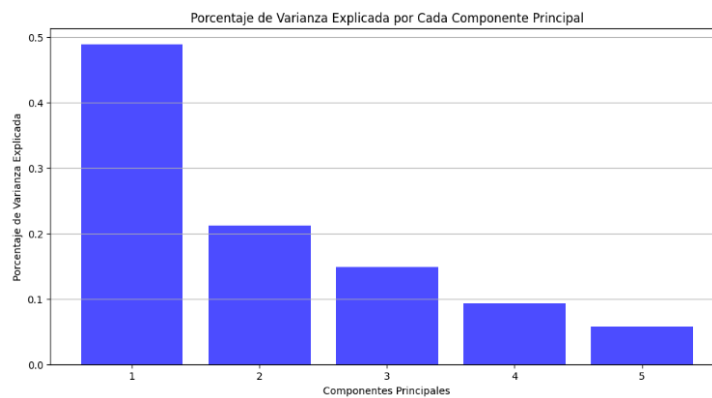
Una vez evaluados los métodos a utilizar, se determinó que, para un modelo de desistimiento, KMeans era preferible debido a su simplicidad y eficiencia, especialmente dado que los datos estaban bien separados y normalizados. Este algoritmo no solo presentó una buena puntuación de silueta, sino que adicionalmente sus clústeres se encontraron bien definidos y agrupados.

De acuerdo con análisis previo del PCA, ver Figura 16, se observó que el 70% de la información se encuentra entre los dos primeros componentes. La decisión de utilizar el PCA y los dos componentes principales se respaldó por los resultados de la varianza, como por las

ventajas prácticas de la reducción de dimensionalidad. Esta elección permite capturar la mayor parte de la información relevante en los datos, facilitando su análisis, interpretación y visualización.

Figura 16

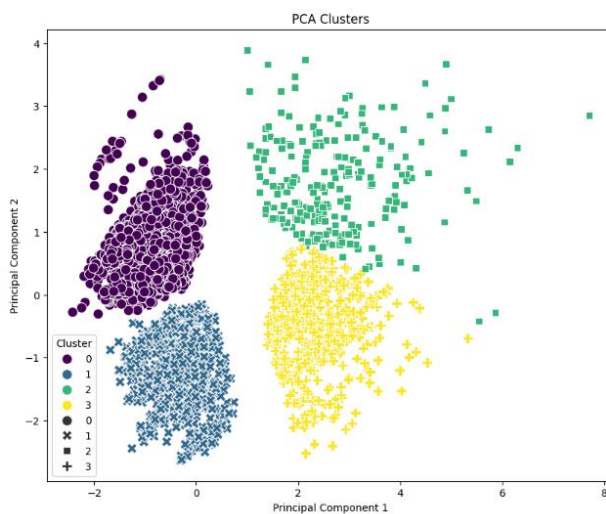
Porcentaje de Varianza de componentes del PCA



En el gráfico de las variables PCA segmentadas en 4 clústeres utilizando el algoritmo de KMeans (Pardeep Kumar, 2012), como se observa en la Figura 17, se identificó las características principales del dataset.

Figura 17

Gráfico PCA clusters



Se identificó y segmentó la información en 4 Clústeres definidos, donde el clúster 0 (Morado) definido y concentrado en la parte superior izquierda sugirió similitudes fuertes entre las observaciones en este grupo, en ellas se destacan que son las personas con más números de hijos, un salario bajo y un porcentaje de recaudación del 55%. Para el clúster 1 (Azul) localizado en la parte inferior se observa un grupo definido, con características como los son salarios más bajos, son los más jóvenes del estudio con una media de 34 años y el porcentaje más bajo de entrega de subsidios con una media del 19%. En el Clúster 2 (Amarillo) se identifica un grupo más disperso, con media de edad de 52 años, un buen porcentaje de recaudación 79% y la media más alta de entrega de subsidios y finalmente el clúster 3 se observan la mayor media de las observaciones con un 83% y el indicador más bajo de hijos. El análisis de medias por clúster se observa en la Tabla 5.

Tabla 5

Identificación de clusters

Variable	Cluster 0	Cluster 1	Clúster 2	Cluster 3
Hijos	1,0220234	0,71414914	0,403361345	0,151802657
Salario	233175,22	169219,25	1049226,24	839218,30
Edad	39,13	34,47	52,87	45,28
Recaudo	0,55	0,48	0,75	0,83
Pagos	0,02	0,00	0,79	0,04
Valor Inmueble	137297235,11	139368059,79	136654351,71	143501963,00
Subsidio	0,94	0,19	0,94	0,69

El uso de KMeans permitió identificar efectivamente a los clientes en cuatro clusters distintos. Esta segmentación facilita la identificación de grupos de clientes con características similares, lo que aportará al modelo una visión estratégica para la retención de clientes por parte de la constructora. KMeans fue seleccionado como el algoritmo más adecuado para el modelo de desistimiento debido a su capacidad para manejar datos globulares, simplicidad y eficiencia.

Entrenamiento Balanceo

Para abordar el problema del desistimiento de clientes en la constructora, se seleccionaron varios modelos de aprendizaje automático, entre ellos Decision Tree, Logistic Regression, Random Forest, KNN, XGB, BAG, ADAR, ADA, META y META2. La elección de estos modelos se basa en su capacidad para manejar diferentes tipos de datos y proporcionar distintas perspectivas sobre el problema, lo que permite una evaluación exhaustiva y precisa.

Estos modelos fueron seleccionados por su robustez y eficacia en problemas de clasificación. Decision Tree y Random Forest, por ejemplo, son conocidos por su capacidad de manejar tanto variables categóricas como continuas y por su interpretación intuitiva. Por otro lado, modelos como Logistic Regression ofrecen simplicidad y eficiencia computacional. La inclusión de modelos como XGB y ADA, que son potentes algoritmos de boosting, asegura que se exploren técnicas avanzadas que pueden mejorar significativamente la precisión predictiva. La diversidad en los modelos seleccionados garantiza que se pueda identificar el enfoque más efectivo y adecuado para predecir el desistimiento de clientes en la constructora. En la Tabla 6 se detalló el comparativo de los indicadores de la evaluación de los modelos seleccionados.

Tabla 6

Comparativo de resultados de modelos

Model	Precisión	AUC	Complejidad	Recall
Decision Tree	44.1	83.76	3	36.95
META	42.5	82.71	6	35.15
XGB	41.3	77.10	6	31.85
BAG	40.7	75.46	7	30.71
KNN	40.1	73.10	6	29.3
ADAR	39.8	69.97	4	27.83
Random Forest	39.7	69.36	6	27.55
META2	39.2	67.07	8	26.25
ADA	39.1	63.34	7	24.74

Logistic Regression	38.4	54.77	2	21.04
----------------------------	------	-------	---	-------

Nota: Esta tabla muestra los resultados de precisión y score para cada modelo probado en el proyecto.

La variable de precisión (accuracy) es una métrica crítica en la evaluación del desempeño de modelos de predicción, especialmente en el contexto del análisis de desistimiento de clientes en una constructora. Esta métrica es esencial para determinar cuán efectivamente el modelo puede identificar tanto a los clientes que van a desistir como a los que no lo harán. Un modelo predictivo con alta precisión permitirá a la constructora identificar correctamente a los clientes con alta probabilidad de desistir, facilitando así la implementación de intervenciones oportunas y personalizadas, como ofertas especiales o seguimientos proactivos. Esto mejora la retención de clientes y optimiza los costos operativos al reducir la cantidad de falsos positivos y negativos, minimizando las intervenciones innecesarias y la pérdida inadvertida de clientes.

Aunque los distintos modelos y algoritmos se usan para problemas de clasificación, el Decisión Tree se destacó como la opción preferida debido a su capacidad para manejar tanto datos continuos como categóricos, lo que lo hace versátil y adaptable, además, los árboles de decisión son modelos fácilmente interpretables, lo que facilita la comprensión de cómo se realizan las predicciones y la integración en el día a día de las empresas. Aunque pueden ser propensos al sobreajuste, se pueden aplicar técnicas de regularización para mitigar este problema y mejorar la generalización del modelo (AMLA, 2023). En la Tabla 7 se detalló las ventajas y desventajas de los modelos evaluados.

Tabla 7*Ventajas y desventajas de modelos evaluados*

Aspecto	Árboles de Decisión	Support Vector Machines (SVM)	Regresión Lasso	AdaBoost
Ventajas	Interpretabilidad alta: Fácil de visualizar y entender. Manejo de datos categóricos y continuos sin preprocesamiento adicional. Captura de relaciones no lineales entre variables.	Potente para clasificaciones con márgenes claros. Robusto ante outliers.	Útil para modelos lineales con regularización. Selección automática de características significativas.	Método de ensemble que mejora la precisión combinando múltiples modelos débiles.
Desventajas	Propenso a sobreajuste si no se poda adecuadamente.	-Complejidad en la implementación y ajuste debido a la necesidad de seleccionar un kernel adecuado y sintonizar varios hiperparámetros. Interpretabilidad limitada en comparación con los árboles de decisión.	Asume una relación lineal entre las variables independientes y la variable dependiente. Puede no capturar toda la variabilidad en casos de interacciones no lineales complejas.	Sensible al ruido en los datos. Menos interpretable que un solo árbol de decisión. Mayor complejidad en términos de implementación y mantenimiento.
Razón para su Elección	Equilibrio óptimo entre interpretabilidad, manejo de diferentes tipos de datos y capacidad para modelar interacciones no lineales. Facilita			

la implementación de estrategias de retención basadas en las predicciones del modelo. Permite actualización y ajuste con nuevos datos cada seis meses.

En contraste, otros modelos asumen una relación lineal entre las variables independientes y la variable dependiente, lo que puede limitar la capacidad para capturar relaciones más complejas en los datos. Por otro lado, Random Forest, aunque ofrece una mayor robustez al combinar múltiples árboles de decisión, también implica una mayor carga computacional en comparación.

Para la predicción y selección del modelo más adecuado, se realizaron pruebas exhaustivas con diversos algoritmos. Finalmente, se eligió el Árbol de Decisión, el cual obtuvo una calificación de 44.1 y un AUC de 83.76. Estos valores indican que el modelo tiene un nivel de predicción sobresaliente y una capacidad significativa para distinguir entre clientes que desisten y los que no. La elección del Árbol de Decisión se justifica no solo por su alta precisión, sino también por su interpretabilidad superior en comparación con otros modelos. Estos indicadores son visualizados en la Tabla 6, la cual destaca la efectividad del modelo al proporcionar claridad y comprensión sobre las decisiones de clasificación.

El análisis de la curva AUC (ver Figura 18), junto con la matriz de confusión (ver Figura 19), confirma que el modelo satisface las necesidades y expectativas planteadas. En el testeo, se encontró que el 79 % de los individuos que desistieron fueron correctamente identificados por el modelo. Es decir, que con la utilización del modelo es 5,6 veces más probable de identificar

correctamente a un cliente que desistirá en comparación con no usarlo. Además, el modelo logró capturar el 44% del total de las personas que efectivamente desisten. Estos resultados son cruciales para la empresa, ya que permiten desarrollar procesos de retención más eficientes y dirigidos.

La ganancia del modelo, es decir, su capacidad para mejorar las estrategias de retención se demuestra al facilitar la identificación de clientes con alta probabilidad de desistimiento. Esta capacidad predictiva del 79% y la identificación del 44% de los desistimientos reflejan la robustez y utilidad del modelo en el contexto de la constructora. Estos datos demuestran que el Árbol de Decisión es una herramienta poderosa para anticipar comportamientos de los clientes y, por ende, diseñar estrategias de intervención más efectivas para mitigar la pérdida de clientes.

Figura 18

Curva ROC árbol de decisión

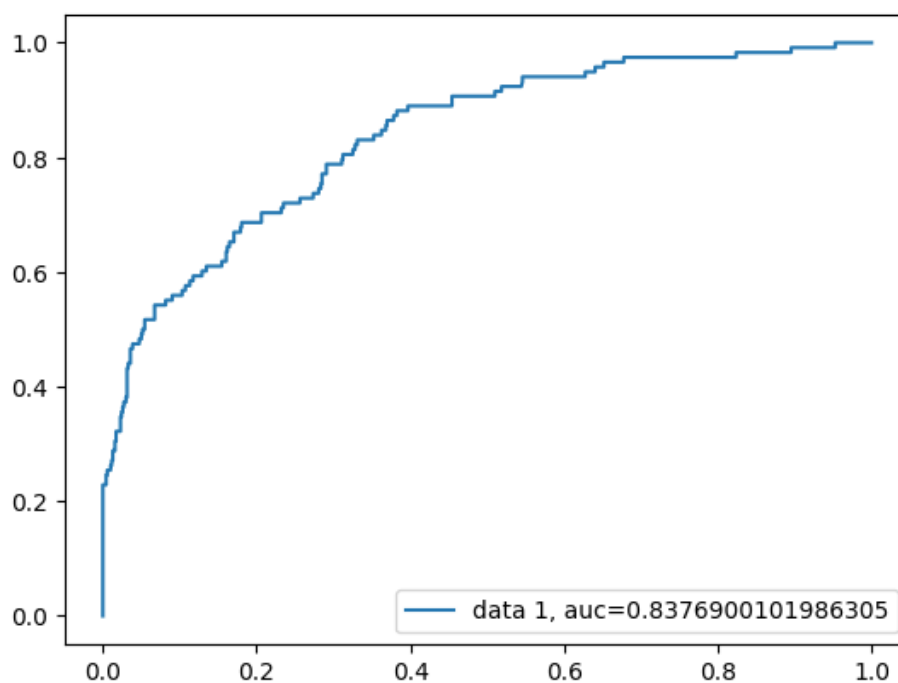
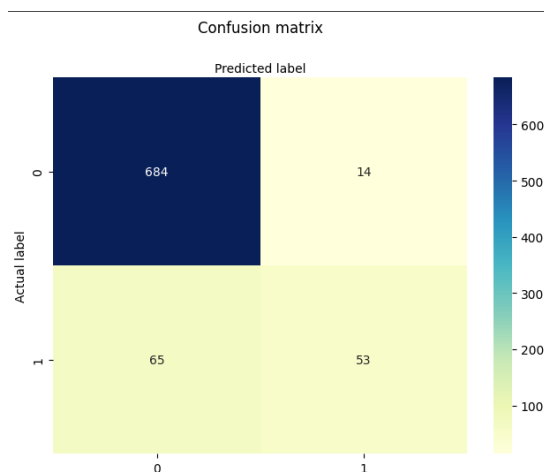


Figura 19

Matriz de confusión



Nota: Cuadrantes de predicción en la matriz de confusión.

Para verificar y validar el modelo, se hace una base de datos “Forecast”, que incluirá las variables necesarias para correr el modelo, y que este nos indique la calificación por cedula.

Modelo Árbol de Decisión

Inicialmente, se procede a cargar los datos de entrenamiento desde un archivo CSV denominado "BASE DESISTIMIENTOS FINAL PYHTON.csv" mediante la función `read_csv()` de Pandas, especificando el delimitador como ";".

Posteriormente, se seleccionan las características relevantes para el entrenamiento del modelo (`X_train`) y la variable objetivo (`y_train`). Estas características incluyen el número de identificación del cliente (CC), el valor del inmueble, el monto del pago, la visita previa (VIS), el salario y la edad del cliente.

Luego, se carga y prepara los datos de predicción desde otro archivo CSV llamado "desistimiento_forecast.csv". Se seleccionan las mismas características que en los datos de entrenamiento para construir `X_forecast`.

Acto seguido, se crea un modelo de árbol de decisión utilizando la clase `DecisionTreeClassifier` de `scikit-learn`. Este modelo se entrena utilizando los datos de entrenamiento (`X_train` y `y_train`) mediante el método `fit()`.

Una vez entrenado el modelo, se utilizan los datos de predicción (`X_forecast`) para predecir las probabilidades de desistimiento para cada cliente mediante el método `predict_proba()` del modelo entrenado. En este caso, se establecen las probabilidades tanto de desistir como de no desistir en 0.5.

Posteriormente, se clasifican las probabilidades de desistimiento en tres categorías: "Bajo", "Medio" y "Alto", utilizando ciertos umbrales predefinidos (0.4 y 0.8). Si la probabilidad es menor que 0.4, se clasifica como "Bajo"; si es mayor o igual a 0.4 y menor que 0.8, se clasifica como "Medio"; de lo contrario, se clasifica como "Alto".

Finalmente, se crea un `DataFrame` que contiene el número de identificación del cliente (CC) y la categoría de desistimiento correspondiente. Los resultados del modelo se imprimen en una lista, mostrando cada cédula con su respectiva clasificación de riesgo de desistimiento, siendo alto, bajo o medio según corresponda. Ver tabla 8.

Tabla 8
Dataset de pruebas

Cédula	Categoría_Desistimiento
1019139828	Medio
659725	Alto
36974521	Alto
102169876	Bajo
52974563	Bajo
1053781795	Bajo
10258284	Bajo
30239992	Bajo
1053797430	Alto
1005568661	Medio
1023952248	Medio

1053769567	Alto
1124860764	Medio
1053853036	Alto
54968236	Alto
13469782	Alto
1093843966	Alto
1023598745	Alto
52268521	Medio
1986634	Alto
102396587	Alto
39556784	Alto
1015465897	Alto

Nota: Resultados de la prueba de predicción

Despliegue Del Modelo

Para el despliegue del modelo, hemos utilizado una base de datos completa que incluye todas las transacciones de ventas realizadas durante el primer trimestre del año 2024. Esta base de datos fue revisada y validada para identificar cualquier patrón o tendencia importante que influyera en la precisión y calidad de la información del modelo.

Después de este proceso, la base de datos fue cargada al modelo para su análisis. Como se muestra en la Figura 20.

Figura 20

Despliegue del modelo

IdVenta	Proyecto	Agrupación	Estado	Comprador Principal Nombre	Comprador Principal Cédula	Vr Recaudo	Vr Agrupacion Incluida la reforma	Fecha Separación	Categoría Desistimiento	%
33568	VESTA ET01	Apto-06-0611	Vendida	RINCON DURAN MILEIDY	21119584	700,000.00	\$ 208,650,000	31/03/2024	Alto	✖
33678	VESTA ET01	Apto-13-1307	Vendida	SALDARRIAGA RINCON PIEDAD CRISTINA	38796050	1,000,000.00	\$ 208,700,000	20/03/2024	Alto	✖
33659	VESTA ET01	Apto-12-1206	Vendida	SALDARRIAGA RINCON PIEDAD CRISTINA	38796050	1,000,000.00	\$ 208,700,000	20/03/2024	Alto	✖
33777	VESTA ET01	Apto-19-1907	Vendida	MORA MOLINA JANE	52864854	1,500,000.00	\$ 208,650,000	30/01/2024	Bajo	✔
33499	VESTA ET01	Apto-04-0411	Vendida	PEREZ PASTRANA INELDO JOSE	73240965	1,400,000.00	\$ 208,650,000	27/02/2024	Medio	⚠
33506	VESTA ET01	Apto-04-0406	Vendida	NARANJO REINA JOSE ELIUMAR	79862168	2,390,000.00	\$ 208,650,000	31/01/2024	Medio	⚠
33762	VESTA ET01	Apto-18-1815	Vendida	MANRIQUE ARISMENDIZ FAVIO NELSON	8829952	700,000.00	\$ 208,650,000	29/02/2024	Bajo	✔
33595	VESTA ET01	Apto-08-0806	Vendida	CALLEJAS BORDA RUBIELA	52132775	1,400,000.00	\$ 208,650,000	24/01/2024	Alto	✖
33469	VESTA ET01	Apto-02-0213	Vendida	CUDRIZ FONTALVO YULIS DEL CARMEN	57085163	1,400,000.00	\$ 201,695,000	15/01/2024	Bajo	✔
33769	VESTA ET01	Apto-19-1905	Vendida	SALINAS FIGUEROA CESAR ESTEBAN	1010069475	700,000.00	\$ 208,650,000	14/02/2024	Bajo	✔
33763	VESTA ET01	Apto-18-1813	Vendida	ROMERO RODRIGUEZ GERALDINE JOHANNA	1022410555	700,000.00	\$ 208,650,000	31/03/2024	Bajo	✔
41964	ALONDRA ET01	Apto-10-103	Vendida	FLOREZ SILVA JENNY XIOMARA	1010233552	1,000,000.00	\$ 208,650,000	27/02/2024	Medio	⚠
41959	ALONDRA ET01	Apto-04-103	Vendida	FORERO RAMIREZ FLAVIO AMADEO	1077083205	5,000,000.00	\$ 208,650,000	18/03/2024	Medio	⚠
40190	ALONDRA ET01	Apto-03-102	Vendida	BARRERA GARZON MARTHA LUCIA	51864094	2,650,000.00	\$ 208,650,000	05/01/2024	Medio	⚠
40754	ALONDRA ET01	Apto-10-102	Vendida	PONTON VALLE ALEXANDER RAFAEL	85466944	1,454,000.00	\$ 208,650,000	19/02/2024	Medio	⚠
41977	ALONDRA ET01	Apto-12-102	Vendida	RAMIREZ GAMBA PAOLA ANDREA	1020814498	1,350,000.00	\$ 208,650,000	27/02/2024	Alto	✖
40210	ALONDRA ET01	Apto-03-602	Vendida	PEREZ TENJO FRANCI NATALIA	1000178400	5,400,000.00	\$ 208,650,000	15/01/2024	Bajo	✔
42064	ALONDRA ET01	Apto-11-504	Vendida	COBOS VIVAS KIMBERLY STEFFANY	1000695588	7,000,000.00	\$ 208,650,000	14/03/2024	Alto	✖
40220	ALONDRA ET01	Apto-11-501	Vendida	GOMEZ LAURA CAMILA	1001175240	2,400,000.00	\$ 208,650,000	31/01/2024	Medio	⚠
40212	ALONDRA ET01	Apto-03-604	Vendida	GARCIA HERRERA JHON STIVEN	1001192319	1,000,000.00	\$ 208,650,000	31/01/2024	Bajo	✔
40219	ALONDRA ET01	Apto-11-502	Vendida	GOMEZ QUIQUE ANYELA MICHEL	1001203745	2,180,000.00	\$ 208,650,000	22/01/2024	Medio	⚠
40734	ALONDRA ET01	Apto-10-602	Vendida	PEÑA SIERRA AURA CAMILA	1001205220	700,000.00	\$ 208,650,000	10/02/2024	Bajo	✔
40736	ALONDRA ET01	Apto-10-504	Vendida	GARCIA GARCIA MARIA FERNANDA	1002331345	2,980,000.00	\$ 208,650,000	22/02/2024	Medio	⚠
42004	ALONDRA ET01	Apto-09-201	Vendida	CUBIDES CASTELLANOS SANTIAGO	1003824657	1,000,000.00	\$ 208,650,000	11/03/2024	Medio	⚠
40224	ALONDRA ET01	Apto-11-401	Vendida	BELTRÁN ROJAS RICHARD ALEXANDER	1010203318	1,599,000.00	\$ 208,650,000	22/01/2024	Bajo	✔
40322	ALONDRA ET01	Apto-04-503	Vendida	LAYTON GALINDO YULY ROCIO	1012348778	2,922,000.00	\$ 208,650,000	25/01/2024	Bajo	✔
42135	ALONDRA ET01	Apto-11-601	Vendida	CAMPOS LIZARAZO KATHERINE	1014228331	1,000,000.00	\$ 208,650,000	30/03/2024	Alto	✖
40326	ALONDRA ET01	Apto-04-603	Vendida	PINZON RUSSI JERSON JAVIER	1015414943	3,400,000.00	\$ 208,650,000	30/01/2024	Bajo	✔
40310	ALONDRA ET01	Apto-04-203	Vendida	MENDEZ FLORIDO BRIGGITTE	1015427346	3,367,000.00	\$ 208,650,000	22/01/2024	Bajo	✔
40324	ALONDRA ET01	Apto-04-601	Vendida	CLAVIJO CAÑAS JOHN EDISON	1018441066	3,700,000.00	\$ 208,650,000	26/01/2024	Bajo	✔
40215	ALONDRA ET01	Apto-11-602	Vendida	RODRIGUEZ RODRIGUEZ JHON ALEXANDER	1018470031	2,650,000.00	\$ 208,650,000	12/01/2024	Alto	✖
39989	ALONDRA ET01	Apto-02-504	Vendida	MARTINEZ AYA DIEGO FERNANDO	1018486872	3,212,000.00	\$ 208,650,000	08/02/2024	Medio	⚠
40201	ALONDRA ET01	Apto-03-401	Vendida	GONZALEZ VALERO NICOLAS	1018499273	1,750,000.00	\$ 208,650,000	22/01/2024	Medio	⚠

Se ha logrado identificar y categorizar el desistimiento en tres niveles dentro del contexto del modelo predictivo. Estos niveles se definen de la siguiente manera:

Bajo: Clientes con una probabilidad baja de desistimiento. Se caracterizan por tener un historial sólido, con una probable estabilidad en sus pagos y una buena relación con la constructora. No presentan indicadores de problemas económicos o socioeconómicos significativos que puedan afectar su continuidad como clientes.

Medio: Clientes con una probabilidad moderada de desistimiento. Aunque muestran ciertos indicadores que podrían sugerir un menor compromiso o interés, aún mantienen una relación estable con la empresa.

Alto: Clientes con una probabilidad elevada de desistimiento. Presentan señales claras de insatisfacción, como retrasos recurrentes en los pagos o quejas frecuentes sobre la calidad de los servicios. Además, pueden enfrentar problemas económicos o socioeconómicos que dificultan su

capacidad para mantener su relación con la constructora. Estos clientes representan un riesgo significativo de abandono y requieren acciones inmediatas para abordar sus preocupaciones y retener su lealtad.

Para facilitar la interpretación y toma de decisiones, se ha generado un icono en Excel que representa visualmente estos niveles de desistimiento. Cada nivel se etiqueta con un icono específico, lo que permite al gestor de negocios centrar su atención en clientes con alta probabilidad de desistimiento, para implementar estrategias de retención efectivas y fortalecer la relación con ellos.

12. Estrategia de Gobierno de datos

Para la gestión adecuada la información, se propone a la constructora implementar un modelo de gobernanza de datos (Vargas, 2019), mejorando así la disponibilidad, integridad y privacidad de las fuentes de información con el objetivo de obtener soluciones analíticas más fiables. Esto optimizará el tiempo en la exploración y preparación de los datos.

De acuerdo con el framework del DAMA-BOK, se establecerá las mejoras estratégicas en cuatro dimensiones importantes para el desarrollo, despliegue y evaluación de los modelos analíticos, estos frentes son, arquitectura, calidad, metadatos, retención, seguridad y políticas, teniendo en cuenta tres pilares fundamentales que son personas, procesos y tecnología. La estrategia global de la gobernanza debe estar enfocada en estos seis frentes principales, como se observa en la Figura 21.

Figura 21*Estrategia de Gobierno de datos*

Actualmente la constructora no cuenta con un área de analítica, ni de gestión de datos, por lo cual, para identificar las estrategias de gobernabilidad de datos sobre los productos analíticos, se realizó el nivel de madurez mediante una entrevista personalizada a un data owner de la compañía. El Data owner seleccionado es una persona que tiene un amplio conocimiento de los procesos, herramientas tecnológicas y de la visión de la compañía. Esta entrevista se dividió en las siguientes áreas de conocimiento:

Administración

Hace referencia a la planificación y gestión de los activos. El objetivo de esta categoría es identificar los responsables de los datos y del manejo de los activos en cada área de negocio, para garantizar las reglas y controles adecuados del manejo de los activos de BI en la compañía.

Administración del riesgo

Se entiende como riesgos a la fuga de información, datos errados, violación a la privacidad y seguridad de los activos. En esta categoría se quiere identificar si existe gestión de los riesgos de los activos dentro de cada área de negocio.

Arquitectura

Hace referencia al diseño de los datos estructurados y no estructurados, aplicaciones que habilita la disponibilidad de los datos y su distribución a los usuarios apropiados dentro de cada área de negocio.

Metadata

Se conoce a la Metadata como una información estructurada sobre datos, que nos los describe sin necesidad de acceder directamente a ellos. En esta Categoría se quiere identificar si existe métodos y las herramientas usadas para crear definiciones semánticas comunes en las áreas de negocio y repositorios de la Metadata.

Auditoria información de Logueo

Se entiende como auditoria de los datos a los procesos para el monitoreo y medición del valor de los datos, mediante el control de los accesos a la información, generación de reportes, borrado de la información entre otros. En esta categoría se quiere identificar si cada área de negocio tiene metodologías para la auditoria de la información utilizada.

Calidad

Se refiere a los métodos para medir, mejorar y certificar la calidad e integridad de los datos. En esta categoría se quiere identificar si existe procedimientos para la calidad de los datos en las áreas de negocio.

Creación de valor

Es el proceso de calificación y cuantificación de los datos para permitir al negocio maximizar el valor creado por los activos. En esta categoría se quiere definir como las áreas del negocio realizan el retorno de la inversión en la recolección, producción y uso de los datos.

Estructura organizacional

Hace referencia a los procesos para el manejo estructurado de los activos, en esta categoría se quiere identificar si cada área de negocio tiene establecida metodologías para el manejo de los datos alineada a las reglas de la organización.

Política

Es la articulación escrita del comportamiento organizacional deseado. En esta categoría se quiere identificar si existe la creación y formalización de políticas y estándares corporativos alrededor del Gobierno de la Información.

Para cada pregunta del área del conocimiento se dan dos tipos de respuesta. ¿Cómo estamos hoy? y ¿Cómo nos vemos en el futuro? Las cuales deben responderse de la siguiente manera: ¿Cómo estamos hoy? V – Verdadero. P–parcialmente. F – Falso Y En ¿Cómo Vemos El Futuro? A - Mantener, M – Mejorar, E I –Invertir, L – Lograr. Se pondero las preguntas del negocio y se dieron un resultado del 1 al 5 en la escala de madurez, según el resultado de la encuesta, como lo muestra la Tabla 9.

Tabla 9

Propuesta de categorización de política gobierno de datos

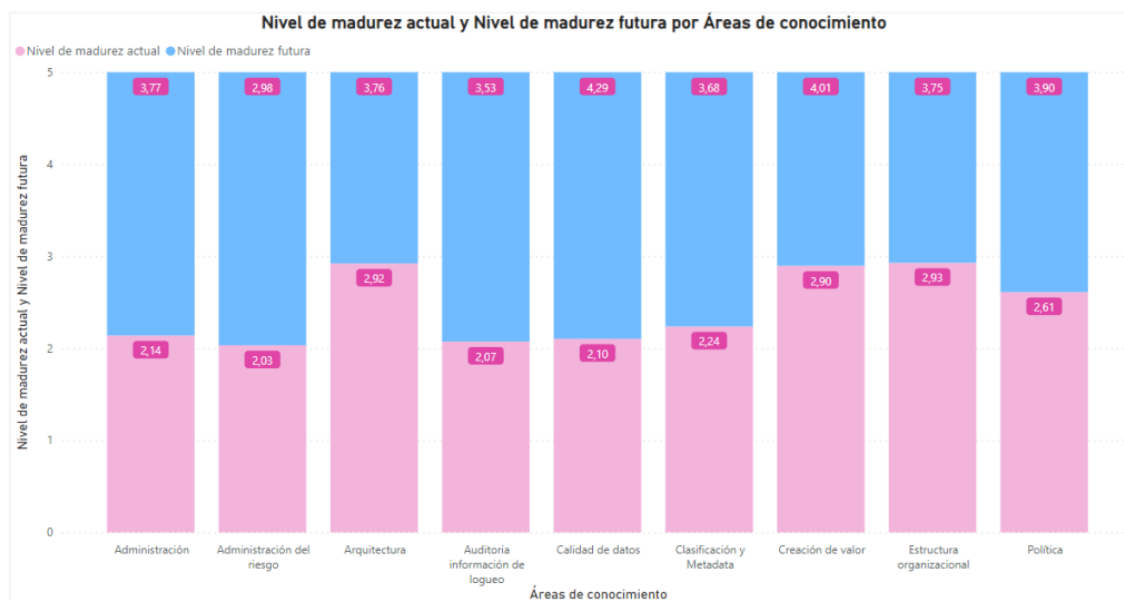
Categoría	Nivel
Nivel 1 Caótico	0 al 1
Nivel 2 Reactivo	1 al 2

Nivel 3 Estable	2 al 3
Nivel 4 Proactivo	2 al 4
Nivel 5 Predictivo	4 al 5

En el Nivel de madurez actual se observó que para el data owner encuestado se tiene falencias en los procesos de administración del riesgo, auditoría información de logueo y calidad de datos, pero a su vez la compañía es más fuerte en la creación de valor y estructura organizacional, ver Figura 22. En el futuro y como ruta estratégica, la compañía debería enfocarse en temas de calidad de datos, siendo la más importante, fortalecer la creación de valor e implementar más políticas de gobernabilidad de datos.

Figura 22

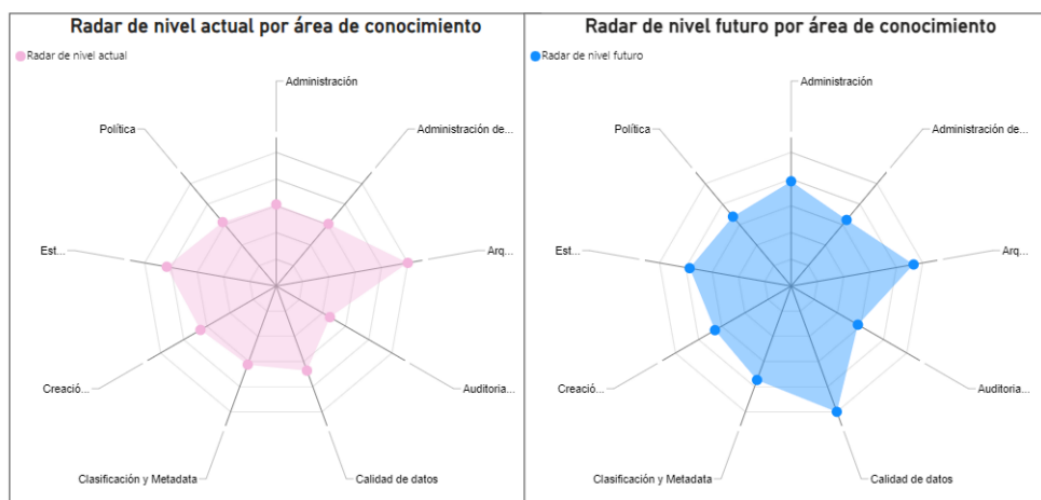
Nivel de madurez de gobierno de datos



Al comparar el nivel de madurez actual y el nivel de madurez futuro se concluye las siguientes estrategias de gestión adecuada de los datos, Ver Figura 23.

Figura 23

Radar de nivel de madurez



El nivel de madurez sobre la gestión de los datos en la compañía es un nivel ESTABLE; lo que quiere decir que en este nivel la organización ha desarrollado un plan de aseguramiento de del valor de los datos y existe compromiso de la alta dirección para orientarse hacia una cultura de la calidad, pero aún no existe una gestión adecuada de la información.

Categorías como Creación de valor, arquitectura y estructura organizacional tienen en promedio los mejores puntajes, ya que la organización considera que se han implementado eficientes estrategias para manejar los activos en la compañía

La compañía, aunque no tiene área de analítica debe mejorar sus estrategias de datos para el otro año y tener una calificación de madurez de 3,36 nivel PROACTIVO, por lo que se recomienda crear el área de análisis clave, procesos automáticos, herramientas de almacenamiento, arquitectura de datos y gobernanza formal dirigido por la alta dirección.

Una vez identificadas las falencias de la compañía., se recomienda ejecutar los siguientes frentes estratégicos para una adecuada gestión de los datos, Ver Tabla 10.

Tabla 10*Modelo de gobernabilidad*

Qué	Cómo
Responsabilidad de Datos	<ul style="list-style-type: none"> • Estructura federada con funciones, comités y sus responsabilidades • Definir los participantes de los datos
Priorización de Datos	<ul style="list-style-type: none"> • Priorización de Datos • Determinar la criticidad de los datos según el negocio
Metadatos y linaje de Datos	<ul style="list-style-type: none"> • Clasificación de metadatos y diccionario de datos • Estructurar el origen de las fuentes de verdad • Crear los elementos de los datos
Calidad y controles de Datos	<ul style="list-style-type: none"> • Dimensiones de calidad de los Datos • Tablero de indicadores de calidad • Crear reglas de calidad de Datos • Supervisar la calidad de los Datos
Seguridad de Datos	<ul style="list-style-type: none"> • Clasificación de seguridad de Datos y privacidad • Definir roles de acceso a la información
Retención de Datos	<ul style="list-style-type: none"> • Contenido de las normativas de retención de Datos • Ciclo de vida de los datos
Políticas, Estándares y Procesos	<ul style="list-style-type: none"> • Crear políticas, estándares y procesos de Gobierno de Datos

Se propone a la constructora adoptar un modelo Central Federado el cual la operación de los datos es responsabilidad de todas las áreas de negocio que administran la información, unificando las prácticas de gestión, obteniendo un centro de excelencia de analítica de datos y es más rápido la escalabilidad de incidencias presentadas en los datos.

La propuesta se enfocó a la creación de Dominios de Información estructurados que sirven para apoyar a los desarrolladores analíticos la gestión de reglas de negocio, comprensión

de los datos y preparación de estos de acuerdo con la ruta de la gobernabilidad propuesta en la Figura 24.

Figura 24

Ruta de gobernabilidad



El proceso para la implementación de dominios de información debe involucrar la colaboración entre el área de Gobierno de Datos, Arquitectura y Business Intelligence. Estos equipos trabajarán en conjunto para estructurar el plan de implementación de dominios y realizarán seguimientos periódicos a este plan.

El área de arquitectura será el encargado de delimitar y articular los activos de datos de acuerdo con los criterios de priorización de los datos y con la política de gobierno de datos. El área de Calidad de datos junto con el responsable del Dominio designado por el comité ejecutivo de Gobierno de Datos verificará la calidad de los datos ingresados y colaborarán en la definición y diseño de los tableros de control necesarios para respaldar la presentación de reportes e informes relacionados con los dominios de información.

Este enfoque garantiza una implementación efectiva, controlada y eficiente de los dominios de datos en la organización.

13. Estimación y Administración de Riesgos

En un entorno tan dinámico como el de la construcción, donde factores externos e internos pueden afectar el desarrollo y la ejecución de proyectos, la gestión de riesgos se convierte en un pilar fundamental para garantizar la viabilidad y el éxito de las operaciones. Como afirmó el autor Tom Kendrick, "la gestión de riesgos no es solo una actividad que realizas una vez y te olvidas de ella, es un proceso continuo que debe integrarse en todas las fases del proyecto".

Desde el inicio de este proyecto, se ha reconocido la importancia de identificar y comprender los riesgos asociados tanto a la investigación como al desarrollo del modelo predictivo. En este sentido, se han diseñado unas estrategias de administración de riesgos que permite abordar proactivamente los posibles obstáculos y maximizar las oportunidades de éxito.

Identificación De Riesgos

En primer lugar, se llevó a cabo una identificación minuciosa de riesgos, centrándose en quienes influyen directamente en la retención de clientes y en la mitigación de desistimientos de negocios. Se analizaron cuidadosamente diversos factores que podrían influir en la precisión y efectividad del modelo, destacando los relacionados con la satisfacción y fidelidad de los clientes. Estos aspectos, aunque no detallados específicamente, son de vital importancia en el contexto operativo y comercial de la constructora.

Este análisis involucra a todas las partes interesadas, desde la alta dirección hasta el personal operativo, así como a clientes, proveedores y contratistas, cuya participación es esencial para abordar de manera efectiva los riesgos identificados. La tabla a continuación detalla los

riesgos específicos que se han identificado y que serán objeto de análisis y planificación estratégica para su mitigación.

Tabla 11

Identificación de riesgos de la constructora

Riesgo	Frecuencia	Causa	Catalizadores	Intangibles
Riesgos económicos o financieros	Media - alta	Cambios en la economía	Recesiones económicas Cambios políticos	Incertidumbre política
		Variación en tasas de interés de créditos de vivienda	Política monetaria restrictiva	Especulación de mercado
		Desempleo del comprador o grupo familiar	Inflación alta	Comportamientos inseguros
		Disminución de ingresos del comprador o grupo familiar	Aumento en el costo de vida	Comportamientos inmorales
		Mal manejo Financiero		Comportamientos desleales
		Cambios normativos en subsidios de vivienda		
		Negación de subsidios o créditos		
		Cambios en las condiciones de los créditos		
Mercado Inmobiliario	Media - alta	Cambios de demanda	Oferta de vivienda	Incertidumbre en el sector
		Cambios en preferencias de vivienda	Cambios en el estilo de vida	Incertidumbre sobre la valoración de las viviendas

		Aumento de competidores en el mercado	Desarrollo de nuevos mercados	
		Ofertas más atractivas de otras viviendas	Cambios en la preferencia del tamaño de la vivienda	
			Beneficios para empresas constructoras	
			Cambios de gobierno	
Cambios en la situación personal del comprador	Media - alta	Muerte del comprador	Recesiones económicas	Comportamientos inseguros
		Enfermedad del comprador	Fluctuaciones en el mercado laboral	Comportamientos inmorales
		Enfermedad familiar	Aumento de la inseguridad	Comportamientos desleales
		Cambio de situación familiar	Matrimonio o divorcio	
		Nuevos integrantes en el núcleo familiar	Desempleo	
		Problemas o conflictos legales		
		Cambios en las prioridades de vida del comprador		
Problemas de construcción	baja - Media	Retrasos en la construcción	Incumplimiento de proveedores	Condiciones climáticas
		Defectos en la construcción	Escasez de materia prima	Desastres naturales
			Procesos de construcción o acabado deficientes	Pandemias y crisis sanitarias
			Desafíos geotécnicos	Fraude o corrupción
			Cambios en el lugar de construcción	Expectativas no cumplidas

			Problemas de mano de obra	
			Problemas de permisos o aprobaciones	
			Cambios de diseños y especificaciones	
			Problemas de financiamiento	
			Publicidad engañosa	
Problemas de financiamiento interno	Media - alta	Falta de capital de trabajo	Problemas de cobranza	Reputación financiera
		Retiro de inversiones	Altos niveles de inventario	Daño a la marca
		Insolvencia	deudas	Pérdida de confianza en el mercado
		Mora de los clientes	Crisis económicas	
			Fluctuaciones en los costos de materias primas	
			Desastres naturales o crisis	

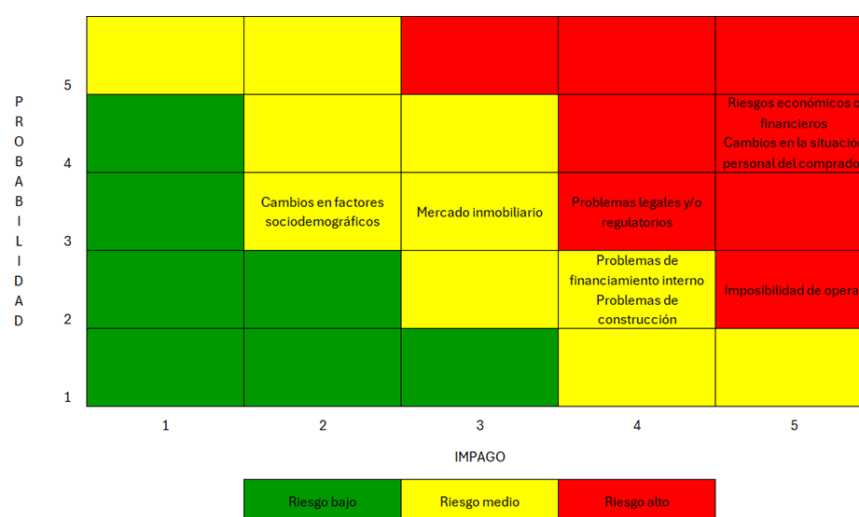
En la Tabla 11 se detalló una amplia variedad de riesgos que la constructora enfrenta en su operación cotidiana, desde riesgos económicos y financieros hasta problemas de construcción y factores sociodemográficos.

Cada riesgo está meticulosamente categorizado según su frecuencia de ocurrencia, posibles causas, catalizadores y factores intangibles asociados. Este análisis exhaustivo ofrece una visión holística de los desafíos potenciales que pueden surgir durante la ejecución de proyectos de construcción.

Al reconocer y comprender estos riesgos, se pueden tomar medidas proactivas para mitigar su impacto y maximizar las posibilidades de éxito. En este sentido, la siguiente etapa de la metodología de administración de riesgos implica la creación de una matriz de probabilidad de riesgo. Esta matriz permite asignar niveles de probabilidad a cada riesgo identificado, evaluando así su impacto potencial en el proyecto, Ver Figura 25.

Figura 25

Niveles de riesgos



Al relacionar los riesgos detallados en la tabla (3) con la matriz de probabilidad de riesgo, se prioriza la atención y recursos en aquellos riesgos que tienen el mayor potencial de afectar negativamente la retención de clientes.

La matriz refleja la necesidad de un enfoque integral de gestión de riesgos que aborde tanto los riesgos más probables como los menos probables, pero de alto impacto. Esto incluye medidas proactivas para mitigar los riesgos financieros, legales y operativos, así como una vigilancia constante del entorno del mercado y la calidad de la construcción.

Estrategias De Mitigación de Riesgos

Después de identificar los riesgos potenciales en el negocio inmobiliario de la constructora a través de la matriz de probabilidad de riesgos, es esencial desarrollar estrategias efectivas de mitigación para cada uno de ellos. Estas estrategias no solo están diseñadas para anticipar y enfrentar posibles contratiempos, sino que también tienen como objetivo principal disminuir significativamente los desistimientos en la compañía.

Al implementar medidas específicas dirigidas a abordar estos riesgos, la constructora no solo asegura la confianza y la satisfacción del cliente, sino que también fortalece su posición en el mercado y garantiza el crecimiento sostenible del negocio. A continuación, se detallarán las estrategias diseñadas para contrarrestar los riesgos asociados con la deserción y el desistimiento de clientes, destacando su importancia para el éxito a largo plazo de la constructora en el mercado inmobiliario. Ver Tabla 12.

Tabla 12

Estrategias de los riesgos

Riesgos económicos o financieros	
Estrategias	
Cambios Económicos	<ul style="list-style-type: none"> • Monitorear constantemente los indicadores económicos para anticipar cambios.
Variaciones en Tasas de Interés	<ul style="list-style-type: none"> • Ofrecer opciones de financiamiento fijo a los compradores. • Establecer acuerdos de cobertura para tasas de interés. • Explorar instrumentos financieros que permitan mitigar el riesgo de variaciones en tasas de interés.
Desempleo	<ul style="list-style-type: none"> • Ofrecer opciones de seguro de desempleo para los compradores, que podrían ayudar a cubrir algunos gastos mientras se emplean nuevamente.

	<ul style="list-style-type: none"> • Realizar evaluaciones exhaustivas de la capacidad de pago antes de firmar un negocio de venta de vivienda, considerando la estabilidad laboral. • Implementar programas de asistencia financiera o flexibilidad en los pagos en casos de disminución temporal de ingresos.
Mal manejo financiero	<ul style="list-style-type: none"> • Ofrecer servicios de asesoramiento financiero para ayudar a los compradores a gestionar sus finanzas de manera responsable. • Implementar programas educativos para compradores sobre gestión financiera y responsabilidad crediticia. • Revisar y actualizar periódicamente las condiciones de los compradores para garantizar que sigan siendo viables.
Subsidios de vivienda	<ul style="list-style-type: none"> • Mantenerse al tanto de los cambios normativos y ajustar estrategias y mantener una comunicación clara con los solicitantes sobre los criterios de elegibilidad y proporcionar orientación para mejorar las posibilidades de aprobación.
Cambios en el contrato de venta	<ul style="list-style-type: none"> • Adoptar una política de precios que sea adaptable a cambios en la oferta y demanda del mercado, así como a condiciones económicas generales. • Establecer asociaciones con desarrolladores, agentes inmobiliarios u otras partes del sector para acceder a nuevas oportunidades y compartir riesgos.

Cambios en la situación personal del comprador

Estrategias

Fallecimiento del comprador	<ul style="list-style-type: none"> • Facilitar la planificación sucesoria para que los herederos puedan gestionar eficientemente la propiedad en caso de fallecimiento.
Enfermedad	<ul style="list-style-type: none"> • Promover la obtención de un seguro de salud integral que pueda cubrir gastos médicos y proporcionar estabilidad financiera en caso de enfermedad. • Implementar opciones de flexibilidad en los pagos de las cuotas para aquellos que enfrentan problemas temporales debido a enfermedades.

Cambios en la situación familiar	<ul style="list-style-type: none"> Realizar evaluaciones continuas de la situación familiar para ajustar las condiciones de los negocios según sea necesario.
Nuevos integrantes en el núcleo familiar	<ul style="list-style-type: none"> Antes de la compra, evaluar las necesidades de espacio futuro para acomodar nuevos integrantes y garantizar que la propiedad sea adecuada a largo plazo. Considerar propiedades con opciones de expansión para adaptarse a cambios en el tamaño de la familia.
Problemas o conflictos legales	<ul style="list-style-type: none"> Ofrecer algún tipo de asesoría a los compradores que les permita obtener herramientas para la solución de sus conflictos
Cambios en las prioridades de vida del comprador	<ul style="list-style-type: none"> Ofrecer asesoramiento personalizado para comprender las prioridades de vida del comprador y adaptar las opciones de vivienda. Incluir cláusulas de flexibilidad en los contratos que permitan ajustes si las prioridades del comprador cambian significativamente

Problemas de financiamiento interno

Estrategias

Capital de trabajo	<ul style="list-style-type: none"> Mantener reservas financieras suficientes para hacer frente a posibles fluctuaciones en los costos operativos y a eventos imprevistos. Realizar análisis periódicos del flujo de efectivo para identificar posibles brechas y tomar medidas preventivas. Establecer líneas de crédito renovables como medida de respaldo para asegurar la continuidad de las operaciones.
Fluctuaciones en el Mercado de Valores:	<ul style="list-style-type: none"> Monitoreo constante de los mercados y ajuste de estrategias financieras. Reserva para compensar la posible disminución en el interés de los inversores. Establecer un sistema de monitoreo constante para detectar señales tempranas de cambios económicos y ajustar estrategias en consecuencia.

Retiro de inversionistas:	<ul style="list-style-type: none"> • Diversificar las fuentes de financiamiento para reducir la dependencia de un solo inversionista o entidad
Mora de los clientes:	<ul style="list-style-type: none"> • Establecer políticas de cobranza claras y comunicarlas de manera efectiva a los clientes desde el principio. • Enviar recordatorios de pagos y mantener una comunicación proactiva con los clientes en riesgo de mora para abordar problemas antes de que se intensifiquen. • Implementar programas de renegociación de deudas para clientes que enfrentan dificultades financieras temporales.
Problemas de construcción e imposibilidad para operar	
Estrategias	
Retrasos en la construcción:	<ul style="list-style-type: none"> • Mantener reservas financieras suficientes para hacer frente a posibles fluctuaciones en los costos operativos y a eventos imprevistos. • Mantener una comunicación transparente y regular con los compradores, informándoles sobre el progreso y cualquier cambio en el cronograma. • Establecer expectativas realistas desde el principio, destacando la posibilidad de retrasos ocasionales y explicando las razones subyacentes. • Incluir cláusulas de compensación en los contratos que ofrezcan incentivos a los compradores en caso de retrasos significativos. • Colaborar estrechamente con contratistas y subcontratistas, asegurando una ejecución eficiente y evitando conflictos que puedan llevar a retrasos.
Defectos en la construcción:	<ul style="list-style-type: none"> • Implementar programas de control de calidad rigurosos en todas las etapas de la construcción, desde la planificación hasta la entrega. • Cumplir con certificaciones y estándares de construcción reconocidos para garantizar la calidad y seguridad de las viviendas. • Ofrecer garantías extendidas que cubran defectos en la construcción por un período significativo después de la entrega.

-
- Fomentar la retroalimentación de los compradores durante y después del proceso de construcción para identificar y abordar cualquier problema rápidamente.

-
- Retención:
- Establecer políticas de retención que ofrezcan seguridad a los compradores en caso de retrasos debido a situaciones externas fuera del control de la empresa.
-

(Saúl Rick Fernández Hurtado, 2019) propuesta aborda de manera exhaustiva las preocupaciones relacionadas con los desistimientos en la compañía. Para contrarrestar los riesgos económicos o financieros, se proponen estrategias como el monitoreo constante de indicadores económicos y la oferta de opciones de financiamiento fijo, asegurando así la estabilidad financiera tanto para la constructora como para sus clientes. Además, se ofrecen soluciones innovadoras, como el asesoramiento financiero personalizado y la implementación de programas educativos sobre gestión financiera, para abordar posibles problemas de mal manejo financiero.

En cuanto a los cambios en la situación personal del comprador, las estrategias incluyen facilitar la planificación sucesoria en caso de fallecimiento y promover opciones de seguro de salud integral para proporcionar estabilidad financiera en casos de enfermedad. Estas medidas no solo ofrecen seguridad a los clientes en momentos difíciles, sino que también fortalecen la relación entre la constructora y sus compradores, fomentando la lealtad y la confianza.

14. Función De Utilidad (VAN) Valor Actual Neto.

El presente estudio ha permitido evaluar la viabilidad financiera de la implementación de un modelo de predicción de desistimientos en una constructora, a través del análisis del Valor

Actual Neto (VAN). Para ello, se consideraron los costos asociados con la adquisición del software, la capacitación del personal y los beneficios netos proyectados durante un período de 5 años.

El Valor Actual Neto (VAN) del proyecto asciende a \$9,918,955.38. Esta cantidad positiva indica que los beneficios netos esperados de la implementación del modelo superan los costos iniciales de \$50,000,000, descontados a una tasa del 10% anual. En la tabla 7 se muestran los cálculos iniciales estimados. Ver Tabla 13.

Tabla 13

Cálculo aproximado de costos iniciales

Costo	Valor del costo inicial
Mejora continua del modelo	\$10.000.000
Actividades de entrenamiento de personal de ventas	\$5.000.000
Computadores y hardware	\$10.000.000
Actualización de bases y calidad de datos	\$10.000.000
Talento Humano especializado	\$15.000.000
Total, costos iniciales	\$50.000.000

En este caso, los "beneficios netos proyectados" se refieren a la diferencia entre los flujos de efectivo netos generados por el proyecto y los costos asociados a la inversión inicial. Se consideran los ingresos estimados que se espera obtener con la implementación del modelo de predicción de desistimientos, descontados los costos iniciales del proyecto. Ver Tabla 14.

Tabla 14

Ingresos iniciales estimados

Beneficio	Estimación del beneficio
Tiempo de ventas por desistimiento	\$1.000.000
Marketing	\$500.000
Documentación y legales	\$1.000.000
Costo por desistimiento	\$2.500.000
Recuperación por desistimiento	\$2.500.000
Razón de éxito esperado	3%
No. Desistidos que iban a desistir por año	8 por año aprox.

La tasa de éxito esperada del 3% y la tasa de descuento del 10%, se derivan de consideraciones específicas relacionadas con el proyecto y el mercado. La tasa de éxito del 3% se refiere a la proporción esperada de clientes que se espera que desistan de la compra de una propiedad en un año determinado, lo que se traduce en un ingreso estimado por desistimiento y su recuperación. Por otro lado, la tasa de descuento del 10% representa el costo de oportunidad de los fondos invertidos, reflejando el rendimiento mínimo esperado por los inversores para compensar el riesgo y la pérdida de liquidez asociados con el proyecto.

$$VAN = \frac{20,000,000}{(1+0.10)^1} + \frac{25,000,000}{(1+0.10)^2} + \frac{30,000,000}{(1+0.10)^3} + \frac{35,000,000}{(1+0.10)^4} + \frac{40,000,000}{(1+0.10)^5} - 50,000,000$$

$$VAN = \frac{20,000,000}{1.10} + \frac{25,000,000}{(1.10)^2} + \frac{30,000,000}{(1.10)^3} + \frac{35,000,000}{(1.10)^4} + \frac{40,000,000}{(1.10)^5} - 50,000,000$$

$$VAN = \frac{20,000,000}{1.10} + \frac{25,000,000}{(1.10)^2} + \frac{30,000,000}{(1.10)^3} + \frac{35,000,000}{(1.10)^4} + \frac{40,000,000}{(1.10)^5} - 50,000,000$$

$$VAN = 18,181,818.18 + 20,661,157.02 + 22,460,142.75 + 24,022,856.28 + 24,592,981.15 - 50,000,000$$

$$VAN = 59,918,955.38 - 50,000,000$$

$$VAN = 9,918,955.38$$

Los flujos de efectivo netos proyectados fueron \$20,000,000 en el año 1, \$25,000,000 en el año 2, \$30,000,000 en el año 3, \$35,000,000 en el año 4, y \$40,000,000 en el año 5. La tasa de descuento utilizada fue del 10% anual, y el costo inicial representa la inversión inicial necesaria para llevar a cabo el proyecto, que incluye la adquisición del software y la capacitación del personal. Que, para esa estimación, el costo inicial fue de \$50,000,000.

Por lo tanto, en términos financieros, el proyecto de implementación del modelo de predicción de desistimientos resulta altamente rentable para la constructora. La obtención de un VAN positivo sugiere que la adopción de esta solución permitirá generar un valor adicional significativo para la empresa a lo largo del período de evaluación.

Este resultado respalda la viabilidad y la conveniencia de avanzar con la implementación del modelo de predicción de desistimientos en la constructora. Además, sienta las bases para tomar decisiones informadas y estratégicas que contribuyan al éxito y la sostenibilidad del negocio en el mercado actualmente competitivo.

15. Conclusiones E Insights

Al realizar toda la investigación y tesis, se concluyen varios aspectos.

Es probable que la empresa no investigara acerca de los temas tratados en este documento, ya que, aunque tienen información importante y algunas características detectadas, no hay un proyecto o evidencia suficiente que muestre la importancia de la información, los datos y los desistimientos, como una prioridad para la empresa.

La implementación de estrategias sólidas de gobernanza de datos es fundamental para garantizar la calidad, confiabilidad y consistencia de los datos utilizados en el desarrollo de modelos predictivos en la constructora. Esta iniciativa permitirá a la organización alcanzar datos de alta calidad para predicciones precisas y consistencia en la aplicación de modelos predictivos.

El modelo de desistimientos funciona con la información obtenida, después de ser limpiada y estandarizada, sin embargo, se le entregan estrategias de calidad de datos y almacenamiento de estos a la empresa, para permitir la mejora continua del modelo.

El modelo de predicción de desistimiento de clientes en la constructora ha demostrado ser extremadamente positivo, ofreciendo una herramienta poderosa para identificar clientes con alta probabilidad de abandono. Mediante el uso de un modelo de Decision Tree, complementado con la evaluación de otros modelos, se ha logrado una alta precisión y una capacidad predictiva significativa, con un AUC de 83.76%. Esta precisión no solo permite a la empresa anticipar posibles deserciones, sino también implementar estrategias preventivas de retención de manera efectiva. La capacidad del modelo para manejar variables clave como la edad, el salario, el número de hijos, la capacidad de pago y el valor del inmueble subraya su robustez y adaptabilidad a los datos específicos de la constructora. En el testeo, se encontró que el 79 % de los individuos que desistían se identificaron correctamente, y el modelo es 5.6 veces más

probable de identificar correctamente a un cliente que desistirá en comparación con no usarlo. El modelo capturó el 44 % de las personas que efectivamente desisten, facilitando el desarrollo de procesos de retención al interior de la organización.

Además, la implementación continua de este modelo proporciona un enfoque dinámico y evolutivo para la retención de clientes. Al actualizarse cada seis meses con nuevos datos de clientes, el modelo se ajusta y mejora continuamente, asegurando que las predicciones se mantengan precisas y relevantes en un entorno en constante cambio. Este enfoque no solo optimiza la gestión de inventarios y los ingresos, sino que también fortalece la competitividad de la constructora en el sector inmobiliario. En conclusión, la adopción de este modelo predictivo no solo es una inversión en tecnología analítica avanzada, sino también una estrategia integral para la mejora continua y la sostenibilidad del negocio. Este modelo, con una calificación de 44.1 y un nivel de precisión del 79%, proporciona una ganancia significativa al mejorar las estrategias de retención y permitir un enfoque más proactivo en la gestión de la relación con los clientes.

16. Recomendaciones Finales

Para garantizar la utilización de la tesis, y que este documento no sea únicamente como un informe más. Se describen estrategias que se consideran fundamentales para llevar a cabo una mejora en todos los aspectos tratados.

Gobernanza de datos o calidad

Desarrollar un programa formal de gobernanza de datos que defina los objetivos, las políticas, los procedimientos y los roles responsables de la gestión de datos en la organización, mejorando significativamente el desarrollo de modelos predictivos para la toma de decisiones según las necesidades puntuales de la constructora.

Modelo

Continuar utilizando el modelo de predicción es una decisión estratégica acertada, ya que ofrece la oportunidad de capitalizar y mejorar continuamente la capacidad predictiva de la empresa. Al implementar estrategias de mejora de datos, como la captura de información más detallada sobre los clientes, la inclusión de nuevas variables relevantes o la optimización de la calidad de los datos existentes, se puede mejorar la precisión y la relevancia de las predicciones de desistimiento. Esto permitirá a la empresa identificar con mayor precisión a los clientes propensos a desistir, lo que a su vez proporcionará oportunidades para implementar medidas preventivas y retentivas de manera proactiva (Merino, 2023). Se espera que estas acciones conduzcan a una reducción significativa del desistimiento y, por lo tanto, a un aumento en los ingresos, incluso hasta en 100 millones.

Además de continuar con el modelo existente, se pueden considerar algunas recomendaciones adicionales para mejorar aún más la capacidad predictiva y el impacto del modelo en la empresa. Esto incluye la evaluación regular del desempeño del modelo y la actualización periódica de los algoritmos y técnicas utilizadas. Asimismo, se sugiere la incorporación de un enfoque de aprendizaje automático continuo, donde el modelo se actualice de forma dinámica a medida que se disponga de nuevos datos. Esto permitirá que el modelo se adapte a los cambios en el comportamiento del cliente y en el entorno empresarial en tiempo real, maximizando así su efectividad y relevancia. Además, se recomienda la colaboración interdisciplinaria entre los equipos de datos, marketing y operaciones para identificar y abordar de manera integral los factores que influyen en el desistimiento, lo que puede generar soluciones más holísticas y efectivas para retener a los clientes y aumentar los ingresos de la empresa.

Fuerza de ventas

Entrenar a la fuerza de ventas para obtener datos más precisos al registrar los compradores, y podrán entender cuáles son los públicos objetivos que deben llegar para disminuir los desistimientos. Identifica posibles falencias en las ventas.

Clientes

Para los clientes con alta probabilidad de desistimiento, la constructora podrá implementar estrategias de retención alto impacto. Por medio de un programa de seguimiento proactivo, donde un representante de servicio al cliente se comuniquen regularmente con el cliente para reconocer cualquier situación difícil con el pago, que permita identificar el desistimiento de forma más rápida evitando costos adicionales y estrategias personalizadas de pagos y problemas.

Para los clientes con probabilidad media de desistimiento, la constructora se enfocará en ofrecer flexibilidad en el proceso de compra. Además, se proporcionará la posibilidad de

personalizar ciertas características de la propiedad o ajustar los términos del contrato según las necesidades específicas del cliente. Además, se podrían ofrecer opciones de financiamiento flexibles, como planes de pago escalonados o períodos de gracia, para reducir la barrera financiera.

El contacto regular por parte del representante de la empresa designado para hacer seguimiento a posibles desistimientos ofrece una serie de beneficios al mantener una comunicación constante permite a la empresa demostrar un compromiso continuo con la satisfacción y el bienestar de sus clientes. Esto se traduce en la capacidad de brindar asistencia personalizada, responder a preguntas o inquietudes, y resolver cualquier problema que pueda surgir durante el proceso de compra.

Además, el contacto frecuente con los clientes potenciales de desistimiento ayuda a la empresa a identificar y abordar rápidamente cualquier problema o inconveniente que pueda surgir, antes de que afecte negativamente la relación con el cliente. Al mantener una comunicación abierta y transparente, la empresa puede demostrar su compromiso con la resolución de problemas y su disposición para adaptarse a las necesidades cambiantes del cliente.

Esta comunicación es fundamental y proporciona a la empresa información valiosa sobre las preferencias, necesidades y comportamientos de los clientes. Esta retroalimentación puede ser utilizada para mejorar los productos y servicios ofrecidos, desarrollar estrategias de marketing más efectivas y fortalecer la relación con los clientes a largo plazo.

Desistimientos

Siguiendo las recomendaciones de clientes, la empresa debe contar con un empleado, y que esté encargado de correr el modelo, y realizar los seguimientos necesarios, y reporte los progresos, se recomienda que sea al gerente comercial. Se sugiera a la constructora minimizar los incumplimientos por parte de la empresa, ya que representan el 22% de todos los desistimientos,

sugerimos crear una campaña de fidelización con proveedores y contratistas que permitan fortalecer el área de abastecimiento.

17. Referencias

- Collins, G. S., & Moons, K. G. (2019). Reporting of artificial intelligence prediction models. *The Lancet*, 393(10181), 1577-1579. [https://doi.org/10.1016/S0140-6736\(19\)30037-6](https://doi.org/10.1016/S0140-6736(19)30037-6)
- Fernández Hurtado, S. R., Martínez Martínez, L. Á., & Ngono Fouda, R. A. (2019). Barreras que dificultan la planeación estratégica en las organizaciones. *Tendencias*, 20(1), 254-279. <https://doi.org/10.22267/rtend.192001.108>
- Grömping, U. (2015). Variable importance in regression models. *Wiley interdisciplinary reviews: Computational statistics*, 7(2), 137-152. <https://doi.org/10.1002/wics.1346>
- Guerrero, D. (1 de Abril de 2023) ¿Por qué la construcción es clave para que el PIB de Colombia sea bueno en 2023?. *Bloomberg Linea*. <https://www.bloomberglinea.com/2023/04/01/por-que-la-construccion-es-clave-para-que-el-pib-de-colombia-sea-bueno-en-2023/>
- Hanslo, R., & Tanner, M. (2020, September). Machine learning models to predict agile methodology adoption. In 2020 15th *Conference on Computer Science and Information Systems* (pp. 697-704). IEEE. <https://doi.org/10.15439/2020F214>
- Jin, D. H., & Kim, H. J. (2018). Integrated understanding of big data, big data analysis, and business intelligence: A case study of logistics. *Sustainability*, 10(10), 3778. <https://doi.org/10.3390/su10103778>
- Kumar, P., Sehgal, V. K., & Chauhan, D. S. (2012). A benchmark to select data mining based classification algorithms for business intelligence and decision support systems. *International Journal of Data Mining & Knowledge Management Process* 2 (5), <https://doi.org/10.5121/ijdkp.2012.2503>
- Lesmes, L. (21 de Junio de 2023) ¿De qué manera se puede terminar el contrato con una constructora?. *El Tiempo*. <https://www.eltiempo.com/economia/finanzas-personales/asi-puede-terminar-un-contrato-con-una-constructora-779393>
- Merino Araujo, I. (2023). *Gestión e implementación de un modelo predictivo para la optimización de los beneficios de una empresa* (Universitat Politècnica de Catalunya). <http://hdl.handle.net/2117/398431>
- Montes, S. (26 de Febrero de 2024). Así es como el sector construcción pone manos a la obra para salir adelante en 2024. *Forbes Colombia*. <https://forbes.co/2024/02/26/negocios/asi-es-como-el-sector-construccion-pone-manos-a-la-obra-para-salir-adelante-en-2024>
- Morandini, M., Coleti, T. A., Oliveira Jr, E., & Corrêa, P. L. P. (2021). Considerations about the efficiency and sufficiency of the utilization of the Scrum methodology: A survey for analyzing results for development teams. *Computer Science Review*, 39, 100314. <https://doi.org/10.1016/j.cosrev.2020.100314>
- Northcott, R. (2017). When are purely predictive models best?. *Disputatio*, 9(47), 631-656.
- Porter, M. E. (1991). *Techniques for Analyzing Industries and Competitors*. 8a ed. Rio de Janeiro: Campus, 49-58. Free Press

- Roy-García, I., Rivas-Ruiz, R., Pérez-Rodríguez, M., & Palacios-Cruz, L. (2019). Correlación: no toda correlación implica causalidad. *Revista Alergia México*, 66(3), 354-360. <https://doi.org/10.29262/ram.v66i3.651>
- Vargas, A. M. C. (2019). El gobierno de datos: un referente entre el gobierno de TI y la inteligencia de negocios. *Revista Científica Ecociencia*, 6(1), 1-19. <https://doi.org/10.21855/ecociencia.61.186>
- Zhao, Y. (2023). Design of a corporate financial crisis prediction model based on improved ABC-RNN+ Bi-LSTM algorithm in the context of sustainable development. *PeerJ Computer Science*, 9, e1287. <https://doi.org/10.7717/peerj-cs.1287>