



Escuela de Administración
Escuela de Ingeniería Ciencia y Tecnología

Maestría en Business Analytics

Modelo de segmentación para crear estrategias de publicidad y venta en el sector inmobiliario
empleando análisis de datos

Presentado por:

John Alexander Hernández Herrera

Bogotá, D.C. 06 de mayo de 2024



Universidad del
Rosario

Escuela de Administración

Escuela de Ingeniería Ciencia y Tecnología

Maestría en Business Analytics

Modelo de segmentación para crear estrategias de publicidad y venta en el sector inmobiliario
empleando análisis de datos

Presentado por:

John Alexander Hernandez Herrera

Bajo la dirección de:

Ph. D. MS.c. Ing. Ferney Alonso Maldonado

Bogotá, D.C. 06 de mayo de 2024

Contenido

Contenido.....	3
Preliminares	5
Agradecimientos	6
Dedicatoria	7
Declaración de originalidad y autonomía	8
Lista de Ilustraciones.....	10
Lista de tablas.....	12
Abreviaturas	14
Glosario	15
Resumen Ejecutivo.....	16
Palabras clave.....	16
Abstract	17
Keywords.....	17
1. Introducción.....	18
2. Objetivos.....	20
2.1. Objetivo General	20
2.2. Objetivo Especifico.....	20
3. Alcance	21
4. Metodología.....	22
5. Cronograma	24
6. Descripción de la Situación organizacional donde se realizará el proyecto (Contexto)26	
Metodología CRISP - DM.....	28
6.1. Entendimiento del negocio.....	30
6.2. Compresión de los Datos.	33
6.3. Preparación de los datos	38
6.4. Modelado.....	43
6.4.1. Segmentación de datos.....	43
6.4.2. Clasificación de genero según el nombre	46

6.4.3.	Correlaciones de Variables	52
6.4.4.	Implantación de K-Means.....	52
6.5.	Evaluación	56
6.6.	Despliegue.....	76
7.	Descripción de las alternativas, estrategias y/o acciones que se toman en el análisis de la solución a la problemática	79
7.1.	Análisis crítico	82
7.2.	Estrategias.....	82
7.2.1.	Estrategias financieras	82
7.2.2.	Estrategias de innovación	83
7.2.3.	Estrategia de mercadeo	83
7.3.	Objetivos a corto plazo	85
8.	Plan y recomendaciones de implementación y aplicación	94
9.	Conclusiones.....	96
	Referencias bibliográficas.....	98

Preliminares

La experiencia del usuario con gran esfuerzo en la atención al cliente es un factor muy importante en la industria inmobiliaria, debido a que, el modelo de negocio se basa en generar ventas de activos de alta calidad. En este sentido, se puede lograr la conversión y retención de clientes si se aplican las estrategias adecuadas. Por lo tanto, este estudio investiga y mide varias variables relacionadas con la experiencia del usuario para sugerir un conjunto de estrategias efectivas. Para lograrlo, este estudio comienza considerando aspectos conceptuales, revisando sistemáticamente la literatura sobre el tema y aplicando herramientas para que los resultados puedan ser utilizados para abordar aspectos de mejora.

Agradecimientos

En primer lugar, deseo expresar mi agradecimiento a la Universidad del Rosario y a la escuela de administración, por el apoyo que ha brindado a este proceso de formación, por la dirección y el rigor que ha facilitado al mismo. Gracias por la confianza ofrecida desde que llegué a esta escuela.

Así mismo, agradezco a mis compañeros de estudio su apoyo personal y humano, quienes han compartido sus ideas, proyectos e ilusiones durante este tiempo, un trabajo de profundización es siempre fruto de ideas, proyectos y esfuerzos previos que corresponden a otras personas y de quienes nos rodean.

Por su orientación y atención a mis consultas sobre metodología, mi agradecimiento especial al PhD Ferney Alonso Maldonado por la información, tiempo facilitado y las sugerencias recibidas de su parte. Finalmente, gracias a los docentes, por su paciencia, dedicación a este proceso formativo, a esas valiosas sugerencias en momentos de duda.

John Alexander Hernández Herrera

Dedicatoria

Dedico mi tesis principalmente a Dios, por darme la fuerza necesaria para culminar esta meta.

A mis padres, por todo su amor y por motivarme a seguir hacia adelante, desde niño, madre así no estes aquí en esta tierra.

También a mis hermanos, por brindarme su apoyo moral en esas noches que tocaba desvelar.

A quienes creyeron en mí y brindaron su apoyo de una forma directa o en silencio; y, finalmente, a los que no creyeron en mí, con su actitud lograron que tomará más impulso.

John Alexander Hernández Herrera

Declaración de originalidad y autonomía

Declaro 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 que he 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.

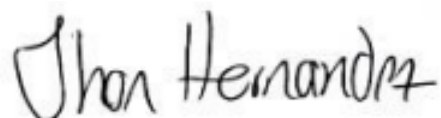


John Alexander Hernández Herrera

Firmado en Bogotá, D.C. el 06 de mayo de 2024

Declaración de exoneración de responsabilidad

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

A handwritten signature in black ink that reads "John Hernandez". The signature is written in a cursive style with a large initial 'J'.

John Alexander Hernández Herrera

Firmado en Bogotá, D.C. el 06 de mayo de 2024

Lista de Ilustraciones

Ilustración 1. Mercados Actuales de la compañía	27
Ilustración 2. Fases de la metodología CRISP - DM	29
Ilustración 3. Organigrama de la compañía	31
Ilustración 4. Dimensiones de base de datos	35
Ilustración 5. Etapas de adquisición de vivienda	37
Ilustración 6. Dimensión BST.....	42
Ilustración 7. Tratamiento de outliers	42
Ilustración 8. Código segmentación edad	44
Ilustración 9. Código segmentación ingresos	45
Ilustración 10. Código segmentación área	46
Ilustración 11. Frecuencia identificación genero primer nombre	49
Ilustración 12. Frecuencia identificación género	51
Ilustración 13. Matriz de correlaciones.....	53
Ilustración 14. Método de codo	54
Ilustración 15. Grafica de clústeres.....	55
Ilustración 16. Grafica de clústeres 3D.....	55

Ilustración 17. Clúster evento compra	57
Ilustración 18. Clúster evento persona.....	59
Ilustración 19. Clúster evento inmueble	60
Ilustración 20. Correlación entre variables	61
Ilustración 21. Grafica de edades.....	62
Ilustración 22. Grafica de ingresos	63
Ilustración 23. Grafica área de inmuebles.....	64
Ilustración 24. Grafica estado civil.....	65
Ilustración 25. Grafica por ciudad	67
Ilustración 26. Grafica del tipo de inmueble.....	68
Ilustración 27. Grafica Ocupación	69
Ilustración 28. Grafica genero.....	70
Ilustración 29. Grafica entidad de crédito.....	72
Ilustración 30. Grafica de fuente.....	74
Ilustración 31. Grafica relación edad y estado civil.....	75
Ilustración 32. Dashboard sin tratamiento de datos	77
Ilustración 33. Dashboard con tratamiento de datos	78
Ilustración 34. Saldo de capital total de la cartera hipotecaria	80
Ilustración 35. Árbol de problemas.....	81

Lista de tablas

Tabla 1. Cronograma de actividades.....	24
Tabla 2. Base de datos nulos.....	34
Tabla 3. Tipo de dato	35
Tabla 4. Estadístico de las variables	36
Tabla 5. Variables del Business Smart Table - BST.....	38
Tabla 6. BST Filtrada.....	40
Tabla 7. Validación de datos nulos	40
Tabla 8. Resumen estadístico.....	41
Tabla 9. Frecuencia segmentación edad	44
Tabla 10. Frecuencia segmentación ingresos.....	45
Tabla 11. Frecuencia segmentación área	46
Tabla 12. División del nombre del comprador	47
Tabla 13. Identificación genero primer nombre	49
Tabla 14. Identificación genero segundo nombre.....	50
Tabla 15. Definición de género.....	51
Tabla 16. Indexación de clústeres	54
Tabla 17. Tipo de evento compra	56
Tabla 18. Clúster evento persona.....	58
Tabla 19. Clúster evento inmueble	59
Tabla 20. Distribución de edad	62
Tabla 21. Rango de ingresos.....	63
Tabla 22. Distribución de área	64

Tabla 23. Estado civil	65
Tabla 24. Distribución por ciudad	66
Tabla 25. Tipo de inmueble	67
Tabla 26. Distribución por Ocupación.....	69
Tabla 27. Distribución de genero.....	70
Tabla 28. Distribución entidad de crédito.....	71
Tabla 29. Distribución de la fuente.....	73
Tabla 30. Actividades de la compañía	86
Tabla 31. Plan de acción estrategia 2.....	88
Tabla 32. Plan de Acción estrategia 3.....	89
Tabla 33. plan de acción estrategia 4.....	91
Tabla 34. Plan de acción estrategia 5.....	92

Abreviaturas

BA: Business Analytics

BDA: Big Data Analytic

VIS: Vivienda de Interés Social

VIP: Vivienda de Interés Prioritario

CRISP-DM: Cross Industry Standard Process for Data Mining

BST: Business Smart Table

Glosario

Análisis de datos: Transformación e interpretación de la información estructurada y no estructurados a través de varias herramientas, técnicas y métodos para descubrir tendencias y resolver problemas.

Clientes: Persona física o jurídica que acepta bienes, servicios de un vendedor a cambio de una forma de pago.

Clúster: Se refiere a un conjunto o grupo de datos con características semejantes, elementos o tributos en común, es utilizado en el análisis de los datos para la identificación de los patrones buscando maximizar la similitud.

K-Means: Es un algoritmo de agrupación no supervisado usado en el análisis de los datos, agrupa los datos en un conjunto de grupos o clústeres donde cada dato pertenece al clúster más cercano en función a la distancia euclidiana.

Mercado: Lugar teórico donde se encuentra la oferta, la demanda de productos y servicios determinado por un precio.

Publicidad: Es una forma de comunicación visual, escrita o auditiva cuya función principal es informar y persuadir a los consumidores sobre productos o servicios, introduciendo nuevas marcas o productos en el mercado de consumo, mejorar la imagen de marca y reposicionar los productos.

Segmentación: La segmentación de mercado es una estrategia de marketing que implica dividir la audiencia de una marca o empresa en pequeños grupos que comparten ciertas características.

Resumen Ejecutivo

Modelo de segmentación para crear estrategias de publicidad y venta en el sector inmobiliario empleando análisis de datos

El objetivo principal de este estudio es diseñar un modelo con implementación tecnológica que permita crear segmentación, estrategias comerciales y personalizadas de los clientes interesados en adquirir vivienda. Asimismo, la compañía está interesado en identificar cual es el género que realizo la mayor compra de inmuebles, teniendo en cuenta que la base de datos no contiene esta información detallada por lo cual se procedió a identificar el género a partir de los nombres de los compradores. En este sentido, para desarrollar este trabajo se utilizó la metodología de Crisp-DM a través de Algoritmo clustering usando el método K-Means, una de las ventajas que tiene este método es que tiene la capacidad de analizar bases de datos con más de diez mil registros. Se pudo concluir que, el análisis en este apartado muestra que la manufactura es una industria con ciertas ventajas competitivas en fortaleza económica, lo que no significa que la empresa sea líder en ciertos cambios, por lo que compañía enmarca la situación actual de lo que sucede en el mercado.

Palabras clave

Segmentación, identificación cliente, análisis de negocios, análisis de datos, sector inmobiliario.

Abstract

Segmentation model to create advertising and sales strategies in the real estate sector using data analysis.

The main objective of this study is to design a model with technological implementation that allows creating segmentation, commercial and personalized strategies of clients interested in acquiring living place. Likewise, the company is interested in identifying which gender made the largest purchase of real estate, taking into account that the database does not contain this detailed information, which is why the gender was identified from the names of the buyers. In this sense, to develop this work the Crisp-DM methodology was used through a clustering algorithm using the K-Means method. One of the advantages of this method is that it has the ability to analyze databases with more than ten thousand records. It was concluded that the analysis in this section shows that manufacturing is an industry with certain competitive advantages in economic strength, which does not mean that the company is a leader in certain changes, which is why the company. It frames the current situation of what is happening in the market.

Keywords

Segmentation, customer identification, business analysis, data analysis, real estate sector.

1. Introducción

En los últimos años en Colombia se ha evidenciado un crecimiento en el área de construcción de inmuebles, este auge se da por el apoyo del gobierno para que las familias colombianas puedan acceder a tener su vivienda propia, por medio de la Vivienda de Interés Social (VIS) y la Vivienda de Interés Prioritario (VIP), lo que ha hecho que las empresas del sector constructor se consoliden en el mercado “El sector de la construcción en Colombia tiene un papel muy importante en el progreso del país y sus actividades repercuten directamente en la economía del mismo”(Pedraza P, 2021).

Este trabajo es la base de la propuesta de solución a una necesidad que nace en la compañía puntualmente en el área de ventas, dentro de los diversos procesos que se llevan en el interior del área, existen una serie de procesos que tienen una importante oportunidad de mejora, lo ideal es que cada proceso se realice de la manera más eficiente permitiendo brindar valores agregados al área, los cuales en ocasiones no se tenían en cuenta y pasaban desapercibidos dentro de la empresa.

La segmentación de clientes es una forma importante de comprender las necesidades y expectativas de estos logrando ofrecerles productos y servicios que satisfagan sus necesidades, es por lo que comprender a sus clientes es crucial para el éxito empresarial y se puede lograr mediante una investigación de mercado. Los autores enfatizan la importancia de la segmentación del mercado y la comprensión del cliente para comprender las oportunidades del mercado y adaptar los productos a los diferentes segmentos del mercado.

En este sentido, se busca mediante los datos históricos que posee la empresa en estudio se analicen y aprovechen al máximo la información recopilada generando herramientas que permitan una adecuada toma de decisiones enmarcando un mayor impacto al momento de implementar estrategias de marketing segmentadas a los clientes potenciales de la constructora.

Es por esto que este estudio tiene como finalidad obtener herramientas que permitan trabajar sus datos “se realiza una clasificación de los algoritmos más usados con la finalidad de encontrar cual otorga una mejor precisión y exactitud” (Alejandra et al., 2021). mediante un trabajo práctico de la maestría Business Analytics de la Universidad del Rosario, donde se realizó un acercamiento con la constructora a través de una alianza empresa – universidad proporcionando la base de los datos e información necesaria para este estudio. Asimismo, la compañía brinda dos funcionarios que realizan seguimiento, propuestas y validación de los avances establecidos para el desarrollo de cada actividad para este estudio.

Por medio de este trabajo se pretende establecer lineamientos para el aprovechamiento de los grandes volúmenes y tipos de datos disponibles en el área de ventas a través de la creación de un marco de trabajo para la implementación de Business Analytics adaptado a las necesidades de la organización, el desarrollo está basado en la investigación de factores que definen esta implementación exitosa. Asimismo, se analizan las variables de estudio por medio de tableros de control.

2. Objetivos

2.1. Objetivo General

Diseñar un modelo con implementación tecnológica que permita crear segmentación, estrategias comerciales y personalizadas de los clientes interesados en adquirir vivienda.

2.2. Objetivo Especifico

- ✓ Identificar las variables demográficas, socioeconómicas y de comportamiento sobre la población objeto de estudio, con el fin de realizar una buena segmentación
- ✓ Conocer e identificar a través de una segmentación las preferencias, deseos, características y gustos en común de los clientes de la compañía.
- ✓ Identificar y analizar el género al que pertenece los compradores de la compañía, para generar campañas de marketing segmentadas.
- ✓ Diseñar clúster de la concentración de los clientes de la compañía sobre los tipos de inmuebles que ofrece la constructora

3. Alcance

El presente proyecto contempla la aplicación de técnicas descriptivas para el desarrollo de un modelo con base a la información disponible históricamente de los clientes de la compañía, este proyecto se desarrollará en un periodo de 18 meses aproximadamente, el proyecto tendrá lugar entre las ciudades donde la constructora tiene actualmente proyectos construidos. A continuación, se establece algunos de los aspectos a trabajar en el marco de la metodología propuesta anteriormente.

4. Metodología

El estudio se desarrollará teniendo en cuenta la metodología de Crisp-DM, la cual guía por procesos cómo podemos desarrollar el trabajo, a través del abordamiento en seis fases; también usaremos la ecuación para la normalización de los datos con la ley de Sturges y el logaritmo de clúster, usando el método K-Means para realizar el análisis y la segmentación de los datos, una de las ventajas que tiene este método es que tiene la capacidad de analizar grandes bases de datos. (Alayon et al., 2020), Otra de las ventajas de este algoritmo del método k-Means es un método sencillo y rápido, las fases de la metodología de Crisp-DM a desarrollar son:

En la primera fase se enmarca la comprensión del negocio: en esta etapa se dará a conocer la compañía, la actividad principal, el mercado que abarcan y la descripción de los problemas que se pretenden solucionar al desarrollar este trabajo.

En la segunda fase es la comprensión de datos: en esta fase analizaremos los datos suministrados por la compañía para objeto de análisis, validaremos la estructura, identificaremos como la compañía almacena y custodia la información y validaremos la composición de los datos.

En la tercera fase se aborda la preparación de los datos: En esta etapa realizaremos los análisis, limpieza, tratamiento de las bases de datos y validaciones de los datos para tener una base de datos mejor consolidada sin outliers para que nuestro modelo realice segmentaciones sin datos atípicos y con los datos normalizados.

En la cuarta fase se realizará el Modelado: el cual generaremos una solución a los problemas planteados por parte de la compañía, esta solución se realizará a través de herramientas que nos permitan analizar los datos.

En la quinta fase Evaluación: se desarrollará una evaluación sobre la implementación del modelo lo cual nos permite evidenciar los aspectos a corregir valorando los resultados iniciales.

En la sexta y última fase Distribución y despliegue: En esta etapa una vez el modelo sea evaluado y aprobado por la compañía se activará el uso de la herramienta desarrollada para el área solicitante la ejecute en el campo de acción.

5. Cronograma

En tabla 1, se muestra el cronograma con la relación de las fases, tareas y actividades a realizar. Este cronograma se estableció por meses, la proyección del periodo de ejecución del proyecto empresarial de 18 meses.

Tabla 1. Cronograma de actividades

TIPO	ACTIVIDAD	TIEMPO (Meses)									
		2	4	6	8	10	12	14	16	18	
Fase 1 (Antecedentes - Revisión Literaria)											
Tarea 1.1	Revisión Bibliográfica										
Fase 2 (Compresión del Negocio)											
Tarea 2.1	Determinar objetivos de negocio										
Tarea 2.2	Inventario de recursos, requisitos, suposiciones y restricciones										
Tarea 2.3	Riesgos y contingencias										
Tarea 2.4	Objetivos de la minería de datos										
Tarea 2.5	Evaluación inicial de herramientas y técnicas										
Fase 3 (Compresión de Datos)											
Tarea 3.1	Recopilación datos iniciales										
Tarea 3.2	Describir los datos recibidos										
Tarea 3.3	Exploración de los datos										
Tarea 3.4	Verificar la calidad de los datos										
Fase 4 (Preparación de los Datos)											

TIPO	ACTIVIDAD	TIEMPO (Meses)							
		2	4	6	8	10			
Tarea 4.1	Seleccionar los datos y justificar la inclusión o exclusión								
Tarea 4.2	Informe de limpieza de datos								
Tarea 4.3	Construir datos, atributos, derivados y registros generados								
Tarea 4.4	Integrar datos combinados								
Fase 5 (Modelado)									
Tarea 5.1	Seleccionar técnicas de modelado								
Tarea 5.2	Generar diseño de prueba								
Tarea 5.3	Construir modelo configurando los parámetros								
Tarea 5.4	Evaluar modelo, parámetros, revisado y ajustes								
Fase 6 (Evaluación)									
Tarea 6.1	Evaluar resultados								
Tarea 6.2	Criterios de éxito empresarial modelo aprobado								
Tarea 6.3	Revisión del proceso								
Tarea 6.4	Lista de acciones posibles								
Fase 7 (Despliegue)									
Tarea 7.1	Plan de empleo								
Tarea 7.2	Monitoreo y plan de mantenimiento								
Tarea 7.3	Reporte final, experiencia documental								
Fase 8 (Documento Proyecto Empresarial)									
Tarea 8.1	Construcción y redacción del documento del proyecto empresarial								

Fuente: Elaboración propia.

6. Descripción de la Situación organizacional donde se realizará el proyecto (Contexto)

En el país las empresas constructoras día a día han venido aumentando la construcción de nuevas VIS y VIP, oficinas, bodegas, áreas comerciales y construcción de ciudades dentro de la ciudad, lo cual conlleva a que sea competitivo cada día el adquirir nuevos clientes para la compra de inmuebles (Cesar Moyano-Alulema, 2020.)

Una de estas empresas constructoras es la compañía, la cual es una empresa colombiana líder en la ejecución de proyectos (Vivienda, oficinas, bodegas, centros comerciales, entre otros) en diferentes zonas del país. En la Ilustración 1, se observa la distribución de los proyectos que se encuentran actualmente, la cual tiene presencia en 19 ciudades.

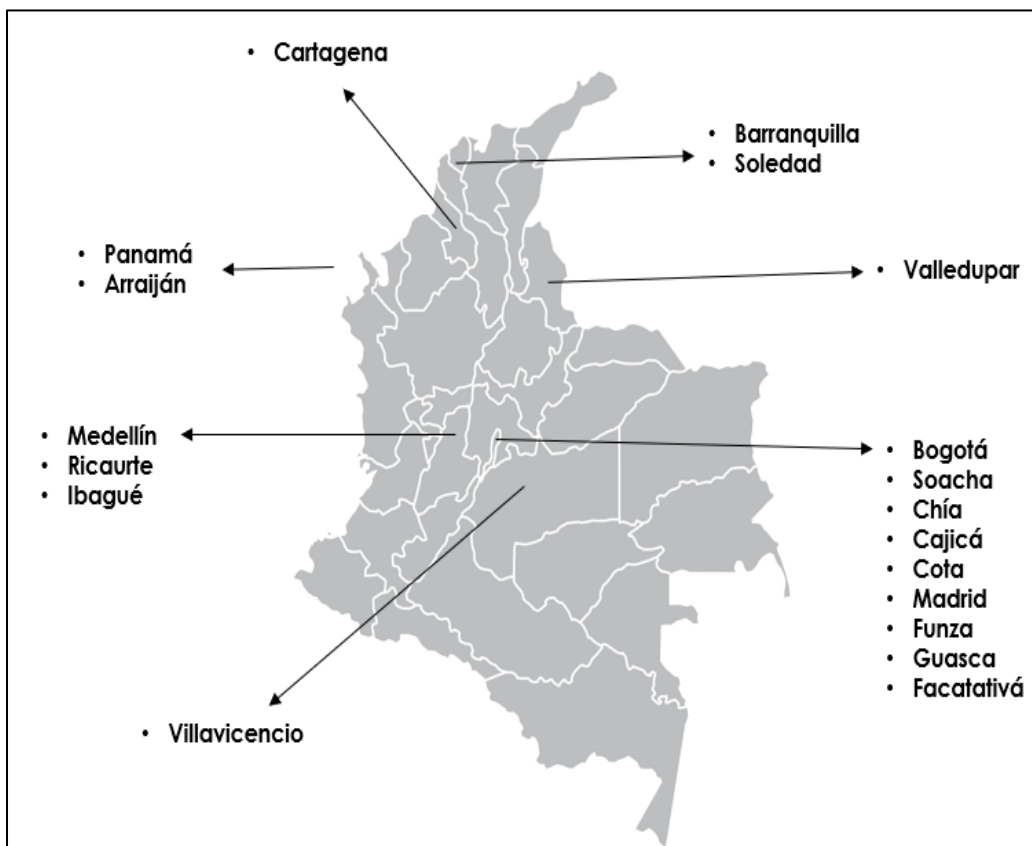
Actualmente, la compañía cuenta con 29 años de experiencia en el sector inmobiliario y aproximadamente con 150 proyectos en ejecución teniendo presencia en Colombia y Panamá. Esta empresa ha revolucionado la industria y el mercado de la construcción y contribuyendo al mejoramiento de la calidad de vida y la realización de los sueños de sus clientes.

El área de marketing de la compañía ha logrado identificar ciertas dificultades para tener acercamiento con sus clientes, comprendiendo los gustos y comportamientos en común; Por lo cual la compañía ha logrado obtener una serie de información¹ estructuradas en ciertas categorías específicas a través de bases de datos (San & Cabrero, 2019). La gran apuesta de la compañía es poder atraer nuevos clientes a través de modelos que permitan identificar los gustos, deseos, objeciones y características personales para realizar campañas mucho más segmentadas y

¹ Algunas de las variables suministradas en las bases de datos son proyecto solicitud, producto, estado, documento del solicitante, nombre, email entre otras.

personalizadas para lograr procesos mucho más eficientes que eviten sobrecostos y pueda generar mensajes de marketing que realmente conecten con los clientes (Alberto et al., 2019.)

Ilustración 1. Mercados Actuales de la compañía



Fuente: Compañía.

De acuerdo a las necesidades antes mencionadas se estableció una alianza entre la compañía y la Universidad del Rosario, por medio de la cual a través de la maestría Business Analytics se acordó desarrollar una solución al problema planteado, a partir de la implementación de un modelo con el cual se pueda estimar con exactitud, efectividad los gustos, deseos de los clientes y el género que realizó la compra del inmueble, aspecto permite la creación de estrategias

de venta y publicidad personalizadas y segmentadas de las personas interesados en adquirir vivienda con la constructora.

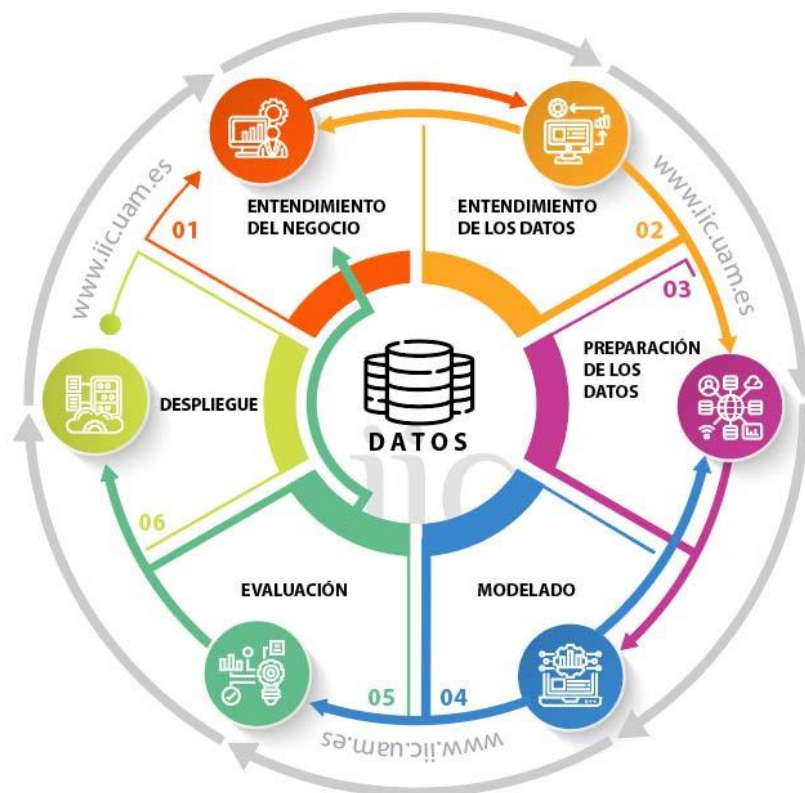
Metodología CRISP - DM

La metodología CRISP – DM es una de las metodologías más usadas en el campo del análisis de los datos y ciencia de datos, esta metodología la usaremos para desarrollar el presente proyecto ya que proporciona una estructura robusta, clara, flexible y bien definida para llevar a cabo la solución a los problemas planteados.

Su enfoque es orientado a resultados, se centra en la entrega de resultados que agreguen valor al negocio, al comprender el negocio y tener definidas las necesidades de los clientes, es una metodología interactiva y cíclica ya que nos permite una mayor flexibilidad y adaptabilidad a la medida que se va avanzado en el proyecto obteniendo nuevos resultados

Como muestra la investigación realizada por expertos en el campo, CRISP-DM puede considerarse la mejor manera de gestionar proyectos dedicados a extraer valor de los datos en las seis fases que se observan en la ilustración 2. En los últimos dos años, el enfoque CRISP-DM ha sido más influyente que otros estándares, como SEMMA de SAS o ASUM-DM de IBM, ya que creó una serie de extensiones CRISP-DM o ejemplos específicos de empresas u tipo de proyecto (Haya Pablo, 2021).

Ilustración 2. Fases de la metodología CRISP - DM



Fuente: (Instituto de Ingeniería del Conocimiento 2023)

Las fases son importantes para comprender lo que está sucediendo y resolver cualquier pregunta que puedan tener sobre el negocio, gracias a su metodología robusta y bien establecida, nos proporciona una guía bien sólida para llevar a cabo paso a paso la solución de los problemas planteados de una manera efectiva y eficiente, convirtiéndose en una herramienta valiosa para el desarrollo de este proyecto.

6.1. Entendimiento del negocio

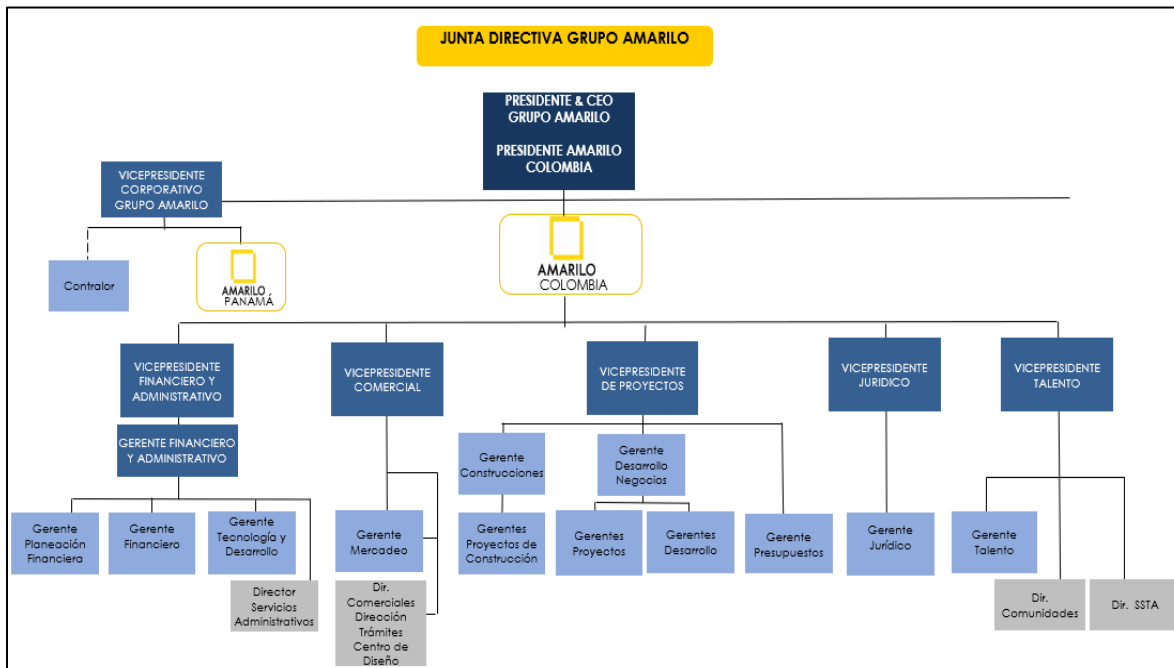
En la primera fase se busca comprender los objetivos y requisitos de la investigación desde una perspectiva comercial, a través del organigrama de la constructora se pretende tener una comprensión de la compañía en la ilustración 3 podemos visualizar como está conformado la compañía, posteriormente convertir este conocimiento en una definición del problema de minería de datos mediante un plan preliminar diseñado para el cumplimiento de los objetivos.

Asimismo, la llegada de la industria 4.0 a los negocios y la relevancia que la transformación digital está tomando, ha llevado, a las organizaciones a implementar diferentes herramientas tecnológicas que les permitan adaptarse rápidamente a los cambios de comportamiento del consumidor y las diferentes disrupciones que se presentan actualmente (Vázquez Rodríguez et al., 2019). Una de esas herramientas tecnológicas que están siendo tendencia en la actualidad, es lo que tiene que ver con el análisis de la información. Los datos se han convertido rápidamente en el activo más valioso que poseen las organizaciones.

Teniendo en cuenta lo anterior, los cargos de gobernanza y analistas de datos están siendo fuertemente demandados por el mercado laboral. El analista de datos juega un papel fundamental en la transformación digital, ya que es un profesional capacitado para aplicar técnicas estadísticas, minería de datos y herramientas computacionales que permitan describir comportamientos emergentes sobre un conjunto masivo de datos² (Ncr & Clinton, 2000).

² Adaptado de <https://www.ucentral.edu.co/noticentral/aumenta-demanda-profesionales-analisis-datos>

Ilustración 3. Organigrama de la compañía



Fuente: Compañía.

La compañía comercializa bienes inmuebles de propiedad raíz, destinados a uso residencial, comercial o empresarial, donde se realiza un proceso de habilitación del suelo, mediante el cual se logran acuerdos de voluntades que permiten la ejecución de proyectos inmobiliarios, con los más altos estándares de calidad, que den respuesta a las necesidades del país.

¿Cuánto facturan en inmuebles vendidos?

Según el DANE en el cierre del año 2022 mediante el informe de sostenibilidad la economía de Colombia creció un 10,6%, cifra que se ubicó dentro del rango de las estimaciones del mercado que esperaban entre el 9,5% y el 2,3% superando las estimaciones del gobierno que preveían un 10.2%. En este sentido, la industria de la construcción impulsó el empleo logrando

ocupar a más de un millón de personas al cierre de octubre de 2021, mientras que, a noviembre, 209.00 nuevos hogares habían alcanzado su sueño de tener vivienda.

¿Porque la compañía quiere hacer este tipo de proyecto?

La compañía con este proyecto quiere es aprovechar al máximo la información que posee de la base de datos, información estructurada en archivos de Excel, la compañía busca con estas bases de datos poder crear programas de fidelización y marketing generando acercamientos a los distintos clientes que actualmente tienen, con un objetivo primordial en la identificación de características y gustos para la adquisición de una nueva vivienda.

Actualmente, con los datos de la compañía se pretende identificar grupos comunes. K-Means es un método de agrupación a distancia que divide los datos en grupos definidos (si hay muchos casos únicos). Los algoritmos basados en distancia se basan en medidas (funciones) de distancia para calcular similitudes entre puntos de datos. Asimismo, busca realizar un análisis demográfico de los estudios adquiridos por sus clientes, permitiendo hacer un clúster acertado por tipo de vivienda adquirida, por rango de precios de las viviendas, rango de edad del comprador mediante visualizaciones de los gustos y características en común, los cuales reúnen un determinado grupo de clientes.

6.2.Compresión de los Datos.

La fase de comprensión de los datos comienza con la exploración inicial de datos, familiarizándose con las variables, identificando si los datos se encuentran estructurados o no estructurados (Kyocera, 2017), “Cuando hablamos de datos estructurados nos referimos a la información que se suele encontrar en la mayoría de las bases de datos” los cuales en ocasiones tienen problemas de calidad, datos incompletos y campos sin información.

Los datos se cargaron a través del entorno colaborativo de Google Colab el cual permite trabajar la información con el lenguaje Python, permitiéndonos poder interpretar y analizar la estructura de los datos cargados en el entorno de Colab utilizando librerías de Python.

Una de las librerías que se uso fue Pandas la cual nos permite realizar la manipulación y el análisis de los datos estructurados, Numpy nos permite realizar cálculos matemáticos matrices, arreglos multidimensionales y vectorizaciones, con Seaborn esta librería está diseñada específicamente para realizar visualización de datos estadísticos atractivos e informativos de fácil interpretación, con Matplotlib crearemos visualizaciones interactivas y animadas en 2D y 3D y para la identificación de genero a partir de los nombres de los compradores utilizamos la librería `gender_guesser.detector`.

La base de datos suministrada por parte de la compañía se encuentra estructurados ya que contienen filas y columnas en un formato estandarizado y definido, la base de consolidado total compradores posee un total de 57 variables, con un total 109.129 filas y 6.220.353 registros de los compradores, su equivalencia de tiempo de información es desde el mes de mayo del año 2006 hasta el mes de agosto del año 2022.

Con la base de datos recibidas por parte de la compañía se procede a revisar su contenido para ver la calidad de esta, ya que se evidencia que ahí información incompleta e incorrecta en su

estructura por lo cual se procederá a hacer un análisis de la información nula para determinar el nivel de limpieza a realizar para las bases y de esta manera al tener las bases más estables se logrará hacer tipos de clústeres de acuerdo con la variable definida(Zuur et al., 2010)

Al cargar la base e imprimir los registros iniciales de los datos sin realizar algún tratamiento el código nos confirma que la base tiene datos nulos campos sin información. En la tabla 2, se puede identificar esta información.

Tabla 2. Base de datos nulos

	Codigo Proyecto	Sala de ventas	Proyecto	Etap	Oportunidad	Ciudad	Agrupacion	Area	Tipo de producto	Tipo de inmueble	...	Fecha llegada Promesa a Trámite	Fecha de prórroga	Fecha de escrituración	Fecha de entrega del inmueble	Fecha notificación inicio firma de promesa	Fecha pactada entrega inmueble
0	NaN	NaN	Las Palmas - Barañ Sub 1 (T 1-2)	NaN	NaN	NaN	APT-1-402	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
1	NaN	NaN	Palladio Sub 1 (T 1,2,3,11,12)	NaN	NaN	NaN	APT-12-506	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
2	NaN	NaN	Cavana Sub 2 (1-9)	NaN	NaN	NaN	APT-3-206	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
3	NaN	NaN	Hacienda Rosablanca Ocarro sub 2 T (6-12)	NaN	NaN	NaN	APT-7-301	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN
4	NaN	NaN	Cavana Sub 1(10-17)	NaN	NaN	NaN	APT-14-206	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN

Fuente: Elaboración propia.

Con la exploración de los datos y las características del dataframe se realiza un resumen del tipo de dato donde 49 variables contienen datos de tipo object y 9 variables poseen el tipo de dato float64 que se refiere a caracteres numéricos con decimales(Marzal Varó & Gracia Luengo, 2009). En la tabla 3, observamos el tipo de dato que contiene el dataframe.

Tabla 3. Tipo de dato

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 109129 entries, 0 to 109128
Data columns (total 57 columns):
#   Column                                     Non-Null Count  Dtype
---  -
0   Codigo Proyecto                          11602 non-null  object
1   Sala de ventas                            11592 non-null  object
2   Proyecto                                  109129 non-null object
3   Etapa                                      15676 non-null  object
4   Oportunidad                               11592 non-null  float64
5   Ciudad                                    11592 non-null  object
6   Agrupacion                                109129 non-null object
7   Area                                       11592 non-null  object
8   Tipo de producto                          11592 non-null  object
9   Tipo de inmueble                          11592 non-null  object
10  Fecha Rad Vta                             11594 non-null  object
11  Fecha Pago Separación / Fecha Comprador  109129 non-null object
12  Fecha de vinculación                       9841 non-null  object
13  Comprador principal                         109129 non-null object
14  Cedula                                      109129 non-null object
15  Fuente                                      11184 non-null  object
16  Comprador Secundario                       1840 non-null  object
17  Cédula comprador Secundario                1840 non-null  object
18  Tipo venta                                 11592 non-null  object
19  Bloq                                        11592 non-null  object
20  Entidad Credito                            10269 non-null object
21  SubTotal                                    11592 non-null  float64
22  Descuento                                  11592 non-null  float64
23  Vr Neto                                    11592 non-null  object
24  Credito                                     0 non-null     float64

```

Fuente: Elaboración propia.

En seguida de consultar el tipo de dato que contiene el dataframe se ejecuta un código para visualizar un resumen estadístico de las variables numéricas de la base datos el cual da el conteo de la información de la variable, la media de la variable, la desviación estándar, el dato mínimo, los tres percentiles del 25%, 50% y 75% y por último el dato máximo que contiene la variable, donde evidenciamos que los datos no se encuentran normalizados tabla 4.

Se consultó el tamaño de la base de datos y su dimensión donde se puede identificar que la base cuenta con 6.220.353 datos y su dimensión es de 109.129 registros distribuidos en 57 columnas o variables, como se observa en la Ilustración 4.

Ilustración 4. Dimensiones de base de datos

```

La base cuenta con : 6220353 datos
La dimensión de la base es de: (109129, 57)

```

Fuente: Elaboración propia.

Tabla 4. Estadístico de las variables

	Oportunidad	SubTotal	Descuento	Credito	Edad	% de participación en la compra	Fecha de desistimiento	Fecha notificación inicio firma de promesa
count	11592	11592	11592	0	11488	11592	0	0
mean	384405	214925294	0	NaN	37	92	NaN	NaN
std	37003	184279638	0	NaN	12	19	NaN	NaN
min	40622	0	0	NaN	0	10	NaN	NaN
25%	367315	140819302	0	NaN	28	100	NaN	NaN
50%	387070	152152000	0	NaN	34	100	NaN	NaN
75%	410626	208690072	0	NaN	44	100	NaN	NaN
max	434339	4496715000	0	NaN	135	100	NaN	NaN

Fuente: Elaboración propia.

Después de consultar el tamaño de la base de datos, se procedió a analizar la cantidad de los datos nulos que se encuentran por cada variable, como se puede observar en el Anexo 2, asimismo, se halló un total de 4.887.176 datos nulos del total de 6.220.353 de la base, el cual representa un 78,56%.

La compañía realiza 5 momentos o etapas sobre los clientes interesados en la adquisición de un inmueble, estos momentos permiten identificar el progreso que lleva el cliente en su adquisición de inmueble las etapas son:

1. **Momento de la visita:** En esta etapa se realiza la captura la información básica del comprador interesado en la adquisición de un inmueble, se procede a dar a conocer los inmuebles disponibles para la venta y se le realiza una proyección de financiamiento para la compra de inmueble.
2. **Momento de la separación:** en esta etapa el comprador muestra un interés en particular por alguno de los inmuebles disponibles que posee la compañía, el cliente informa el

método de financiación para la adquisición del inmueble, el plazo de pago y los recursos financieros que utilizara para la adquisición y se procede a realizar el pago de la cuota de separación del inmueble.

3. **Momento de promesa de compraventa:** en esta etapa se firma un contrato el cual asegura por parte de la compañía separar el inmueble y no ofrecerlo más como un inmueble disponible para futuros compradores interesados en la búsqueda de algún inmueble, de igual manera el cliente interesado, firma y se compromete a llevar a cabo hasta el final de la entrega del inmueble, el pago acordado en las fechas y montos, a través del medio seleccionado para la adquisición.
4. **Momento firma de escritura:** en esta etapa el comprador ya ha cancelado lo correspondiente a la cuota inicial para la adquisición del inmueble por lo cual se procede a la firma de la escritura en las fechas acordadas para que el inmueble pase a ser propiedad del cliente.
5. **Momento entrega:** en esta etapa la compañía hace la entrega oficial del inmueble al encontrarse pago la totalidad del valor, para que sea disfrutado por el cliente que adquirió el inmueble. Ilustración 5. etapas de adquisición de vivienda.

Ilustración 5. Etapas de adquisición de vivienda



Fuente: Elaboración propia.

Teniendo en cuenta la base de compradores después de realizar el análisis anterior se puede concluir que el porcentaje de los datos nulos que nos proyecta con el código, (Ghazzawi & Alharbi, 2019) se debe a que la información está incompleta con datos nulos en los campos, con fecha desde el año 2006 hasta el año 2021; Solo se encontrándose completos en su totalidad, diligenciado y sin datos nulos el año 2022, por lo anterior el porcentaje de los nulos es significativamente alto.

Teniendo en cuenta la base entregada por la compañía se pretende analizar el patrón de consumo, sobre los diferentes proyectos que tiene la compañía, se quiere establecer un rango de precios de los inmuebles y su comportamiento de acuerdo con la ubicación geográfica de los proyectos(Sociedad et al., 2016).

6.3. Preparación de los datos

En la preparación de los datos, se realiza el proceso de limpieza, organización y transformación de los datos, en un formato que pueda analizarse y utilizarse para obtener inteligencia empresarial. Existen múltiples procesos, como limpieza de datos, integración de datos, transformación de datos y mejora de datos, en otras palabras, la preparación de datos es el proceso de convertir datos en un formato que pueda usarse para el análisis(Gamarra, 2020).

La preparación de los datos es una de las etapas de más complejidad en cualquier proyecto de análisis de datos o minería de datos, esta preparación de datos la realizamos para garantizar que los datos estén limpios, completos y listos para nuestro análisis a realizar. Los conjuntos de datos pueden contener errores, valores atípicos, valores faltantes u otro problema que puede afectar la precisión y la validez de los análisis, la preparación incluye la identificación y corrección de estos problemas para garantizar la integridad de los datos ver tabla 5.

Tabla 5. Variables del Business Smart Table - BST

```
[ ] 6.2 # Imprimimos el nuevo business smart table que es nuestro nuevo dataset
```

```
bst.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 109129 entries, 0 to 109128
Data columns (total 16 columns):
#   Column                Non-Null Count  Dtype
---  ---                ---
0   Area                   11592 non-null  object
1   Comprador principal    109129 non-null  object
2   Agrupacion            109129 non-null  object
3   Ciudad                 11592 non-null  object
4   Tipo de producto      11592 non-null  object
5   Ingresos               11458 non-null  object
6   Fuente                 11184 non-null  object
7   Tipo de inmueble      11592 non-null  object
8   Tipo venta            11592 non-null  object
9   Entidad Credito       10269 non-null  object
10  Edad                   11488 non-null  float64
11  Estado civil           11525 non-null  object
12  Ocupación              11459 non-null  object
13  Destino                 11244 non-null  object
14  Vr Neto                11592 non-null  object
15  Vendedor               11592 non-null  object
dtypes: float64(1), object(15)
memory usage: 13.3+ MB
```

Fuente: Elaboración propia.

Los datos necesitan ser transformados o reformados antes de ser utilizados en los análisis, esto puede incluir normalización de los datos, la codificación de variables categóricas, la creación de nuevas características derivadas o la reducción de la dimensionalidad, para esta transformación de los datos, fue aplicada en algunas variables de nuestro modelo con la técnica de la ley de Sturges, ya que nos permite identificar el número adecuado de los intervalos y su amplitud en rangos, lo cual nos permite aproximarnos a una distribución normal.

En ocasiones necesitamos realizar integración de los datos, esto se da porque el origen de los datos proviene de múltiples fuentes y deben integrarse en un solo conjunto coherente, compatibles y consistentes. De igual manera podemos involucrar la selección de las variables más relevantes o significativas para nuestro análisis, esto puede ayudar a reducir la dimensionalidad de los datos y mejorar la eficiencia y la precisión del modelo de análisis como lo podemos observar en la tabla 6 BST filtrada con la preparación de los datos.

Tabla 6. BST Filtrada

```
bst_filtrada.info()

<class 'pandas.core.frame.DataFrame'>
Int64Index: 9884 entries, 97537 to 109128
Data columns (total 16 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Area                 9884 non-null   float64
1   Comprador principal  9884 non-null   object
2   Agrupacion           9884 non-null   object
3   Ciudad               9884 non-null   object
4   Tipo de producto     9884 non-null   object
5   Ingresos             9884 non-null   int64
6   Fuente               9884 non-null   object
7   Tipo de inmueble     9884 non-null   object
8   Tipo venta           9884 non-null   object
9   Entidad Credito     9884 non-null   object
10  Edad                 9884 non-null   float64
11  Estado civil         9884 non-null   object
12  Ocupación            9884 non-null   object
13  Destino              9884 non-null   object
14  Vr Neto              9884 non-null   int32
15  Vendedor             9884 non-null   object
dtypes: float64(2), int32(1), int64(1), object(12)
memory usage: 1.2+ MB
```

Fuente: Elaboración propia.

Los valores faltantes son comunes en los conjuntos de datos y pueden afectar la validez de los análisis, la gestión de esos valores faltantes se realiza mediante técnicas como la imputación de valores, eliminación de datos nulos o variables con valores faltantes excesivos, es importante realizar esta validación para garantizar que los datos cumplan con los criterios de calidad como observamos en la tabla 7 Validación de datos nulos.

Tabla 7. Validación de datos nulos

```
[ ] 6.6 # contamos los nulls de cada columna

bst_filtrada.isnull().sum()

Area                0
Comprador principal 0
Agrupacion          0
Ciudad              0
Tipo de producto    0
Ingresos            0
Fuente              0
Tipo de inmueble    0
Tipo venta          0
Entidad Credito     0
Edad                0
Estado civil        0
Ocupación           0
Destino             0
Vr Neto             0
Vendedor            0
dtype: int64
```

Fuente: Elaboración propia.

Realizamos las verificaciones correspondientes para validar la consistencia de los datos, la identificación de errores y la confirmación de que los datos están en el formato correcto para nuestro análisis, esto los podemos realizar a través de un resumen estadístico de las variables con la función de pandas describe(), permitiéndonos tener una visión rápida del conjunto de datos visualizadas en la Tabla 8.

Tabla 8. Resumen estadístico

```
[25] 6.10 # Consultamos el nuevo resumen estadístico de las variables
bst_filtrada.describe()
```

	Area	Ingresos	Edad	Vr Neto
count	8276	8276	8276	8276
mean	55	3296970	35	176717352
std	11	2202272	11	95742650
min	24	800000	18	30462900
25%	48	1800000	27	138000000
50%	53	2500000	32	150699450
75%	57	4000000	41	165375000
max	117	10000000	76	895935881

Fuente: Elaboración propia.

El tratamiento de los datos outliers o valores atípicos es de gran importancia ya que puede sesgar nuestro modelo, son observaciones que numéricamente difieren significativamente del resto de los datos de la muestra y pueden afectar la media o la desviación estándar, pudiendo llevarnos a estimaciones inexactas, las variables que tratamos con outliers la podemos evidenciar en la ilustración 7 tratamiento de outliers.

Una vez realizada toda nuestra limpieza de datos, transformación y modificación pasamos de tener una base de datos inicial de 6.220.353 datos con 109.129 filas y 57 variables, a tener una base de datos de 132.416 datos con 8.276 y 16 , ilustración 6 Dimensión BST.

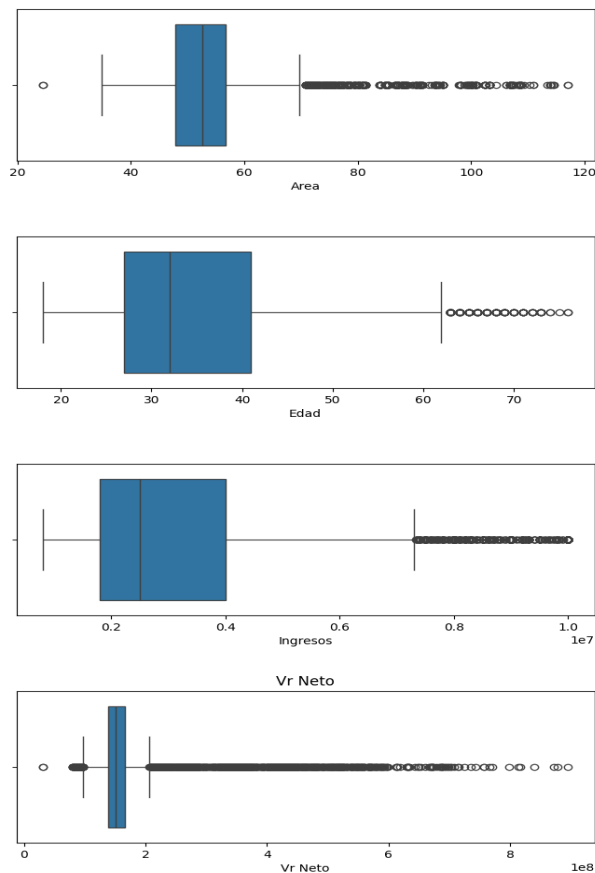
Ilustración 6. Dimensión BST

```
[27] 6.5 # Consultamos el tamaño de la base de datos y su dimension
print("La base cuenta con :", bst_filtrada.size, "datos")
print("La dimensión de la base es de: ",bst_filtrada.shape)

La base cuenta con : 132416 datos
La dimensión de la base es de: (8276, 16)
```

Fuente: Elaboración propia.

Ilustración 7. Tratamiento de outliers



Fuente: Elaboración propia.

6.4. Modelado

En la fase de Modelado es una de las etapas centrales de nuestro proceso, durante esta etapa construimos el modelo de análisis de los datos, con el objetivo de cumplir y resolver los desafíos establecidos en la fase de comprensión del negocio. Por lo cual se selecciona la técnica adecuada que implementaremos para abordar el proyecto y resolver el problema de negocio identificado.

Seleccionaremos las variables predictoras que se usaran para entrenar el modelo, siendo las variables más relevantes teniendo la capacidad de predecir la variable objetivo. En ocasiones dividimos los datos en conjuntos de entrenamiento y prueba, donde el conjunto de datos de entrenamiento se utiliza para entrenar el modelo, mientras el conjunto de prueba se utiliza para evaluar su rendimiento.

6.4.1. Segmentación de datos

En este proceso dividimos un conjunto de datos en grupos o segmentos más pequeños con características similares, esto nos permite entender mejor la estructura de los datos, identificar patrones o tendencias y poder tomar decisiones más informadas basadas en las observaciones encontradas.

Por lo cual, a través de la ley de Sturges, principalmente la utilizamos para determinar el número óptimo de clases, lo que nos puede ayudar en la comprensión de la distribución de los datos, esto nos ayuda determinar cuántos intervalos son necesarios para clasificar un conjunto de datos.

Para las siguientes variables de Edad, Ingresos y Área aplicamos lo anteriormente mencionado de la ley de Sturges donde, $C = 1 + \text{Log}^2(N)$, donde C es el número de clases y N es el número de muestra en el conjunto de datos. Siendo $C = 1 + \text{Log}^{10}(8.276)$; entonces el número de clases es de 5; Una vez identifiquemos el número de clase hallaremos la amplitud de esa clase,

por lo cual la fórmula para calcular la amplitud de la clase es el valor máximo del conjunto de datos, menos el valor mínimo del conjunto de datos sobre el número de clases, dándonos como resultado el rango de las siguientes segmentaciones a partir del código aplicado como lo podemos evidenciar en la ilustración 8, ilustración 9 e ilustración 10.

Una vez aplicado el código realizamos una consulta de la frecuencia de cada una de las variables de edad, ingresos y área, para poder visualizar los resultados observados en la tabla 9, tabla 10 y tabla 11, esto nos permite ver el número de veces que se repite ese dato dentro de la clase que se segmentó.

Ilustración 8. Código segmentación edad

```
[ ] # 7.1 segmentamos la edad de los compradores segun la siguiente escala

def edad (x):
    if x < 18 :
        return "Niño"
    elif x >17 and x < 26 :
        return "Joven"
    elif x > 25 and x < 36:
        return "Adulto Joven"
    elif x > 35 and x< 46:
        return "Adulto"
    elif x > 45 and x< 56:
        return "Adulto Maduro"
    else :
        return "Adulto Mayor"
```

Fuente: Elaboración propia.

Tabla 9. Frecuencia segmentación edad

```
[ ] 7.2 # Sacamos una frecuencia de la variable edad

bst_filtrada.Edad.value_counts()

Adulto Joven    3544
Adulto          1794
Joven           1508
Adulto Maduro   786
Adulto Mayor    644
Name: Edad, dtype: int64
```

Fuente: Elaboración propia.

Ilustración 9. Código segmentación ingresos

```
[ ] # 7.4 segmentamos los ingresos de los compradores por salarios minimos legales vigentes

def ingresos (x):
    salario=1000000
    if x <= salario*2 :
        return "0-2 salarios"
    elif x >salario*2 and x <= salario*4 :
        return "2-4 salario"
    elif x > salario*4 and x <= salario*6:
        return "4-6 salarios"
    elif x > salario*6 and x <= salario*8:
        return "6-8 salarios"
    else :
        return "Mayor a 8 salarios"
```

Fuente: Elaboración propia.

Tabla 10. Frecuencia segmentación ingresos

```
[ ] 7.5 # Sacamos una frecuencia de la variable Ingresos

bst_filtrada.Ingreso_cat.value_counts()

0-2 salarios          3605
2-4 salario           2749
4-6 salarios          876
6-8 salarios          611
Mayor a 8 salarios    435
Name: Ingreso_cat, dtype: int64
```

Fuente: Elaboración propia.

Ilustración 10. Código segmentación área

```
[ ] # 7.6 segmentamos el area de los inmuebles

def area (x):
    if x <30 :
        return "0 a 30 mts cuadrados"
    elif x >30 and x < 46 :
        return "30 a 45 mts cuadrados"
    elif x >45 and x < 61 :
        return "45 a 60 mts cuadrados"
    elif x >60 and x < 76:
        return "60 a 75 mts cuadrados"
    elif x >75 and x < 91:
        return "75 a 90 mts cuadrados"
    else :
        return "Mayor a 90 mts cuadrados"
```

Fuente: Elaboración propia.

Tabla 11. Frecuencia segmentación área

```
[ ] 7.7 # Sacamos una frecuencia de la variable Area

bst_filtrada.Area.value_counts()

45 a 60 mts cuadrados      5735
30 a 45 mts cuadrados      1530
60 a 75 mts cuadrados       538
75 a 90 mts cuadrados       300
Mayor a 90 mts cuadrados    171
0 a 30 mts cuadrados         2
Name: Area, dtype: int64
```

Fuente: Elaboración propia.

6.4.2. Clasificación de genero según el nombre

Uno de los objetivos de identificar el género del comprador a partir de los nombres, fue un gran desafío, ya que requerimos consultar la manera en que podíamos obtener esta característica a partir de los nombres de los compradores.

Para poder determinar el género necesitamos separar el nombre completo que se encontraba concatenado en una misma casilla de información, esto lo realizamos a través de la función `str.split`, este nos permite dividir cadenas en partes más pequeñas basadas por un delimitador específico que se le asigne y devuelve una lista con los nombres y los apellidos de los compradores como lo observamos en la tabla 12. División del nombre del comprador.

Una vez tengamos los nombres separados importamos la librería `request` y `json` estas librerías nos permiten importar módulos externos del código actual, permitiéndonos realizar solicitudes HTTP a una API en línea para trabajar con los datos. Con la biblioteca de Python `gender_guesser`, podemos predecir el género de un nombre dado basándose en estadísticas de nombres de diferentes culturas y regiones a nivel mundial.

Tabla 12. División del nombre del comprador

```

8.1 # Desconcatenamos el nombre del comprador principal
df3=bst_filtrada['Comprador principal'].str.split(" ", expand = True)
# print(df3.iloc[0,0])
df3

```

	0	1	2	3	4	5	6	7
97537	ANGIEE	JULIANA	MARTINEZ	RUEDA	None	None	None	None
97538	NAOMI	JULIETH	CABELLO	MONTERO	None	None	None	None
97540	ANDRES	ERNESTO	GUERRERO	ACOSTA	None	None	None	None
97541	DANIELA	VILLAFANE	RINCON	None	None	None	None	None
97542	RICHARD	SMEINER	HERNANDEZ	RUBIO	None	None	None	None
...
109119	JORGE	FERNANDO	RODRIGUEZ	GRACIA	None	None	None	None
109120	INGRID	JOHANNA	URBANO	MARIN	None	None	None	None
109122	MICHELLE	ROSANA	MENDOZA	RUBIO	None	None	None	None
109123	KAREN	MARGARITA	MOLINA	ROMERO	None	None	None	None
109128	FRANCISCO	JAVIER	HERNANDEZ	VEGA	None	None	None	None

8276 rows x 8 columns

Fuente: Elaboración propia.

La biblioteca `gender_guesser` nos ofrece métodos para poder determinar el género asociado a un nombre, devolviendo cinco posibles resultados del análisis de identificación de género a través de la API, los posibles resultados que nos puede dar son los siguientes:

1. Male = Masculino
2. Female = Femenino
3. Mostly_male = Posiblemente masculino
4. Mostly_female = Posiblemente femenino
5. Unknown = Desconocido

Una vez determinamos el género aplicando la librería con la `api.genderize.io` sobre el primer nombre de los compradores para determinar el género al que pertenecen, podemos ver el resultado que nos generó en la tabla 13. Identificación género primer nombre, la frecuencia del resultado del género a partir del primer nombre lo evidenciamos en la ilustración 11. Frecuencia identificación género primer nombre.

Tabla 13. Identificación genero primer nombre

```
[ ] 8.7 # determinamos el sexo guiandonos por el primer nombre
df3['sexo'] = df3['primer'].apply(str.capitalize).map(lambda x: d.get_gender(x))
df3.head(10)
```

	primer	segundo	pa	sa	a	b	c	d	sexo
97537	ANGIEE	JULIANA	MARTINEZ	RUEDA	None	None	None	None	unknown
97538	NAOMI	JULIETH	CABELLO	MONTERO	None	None	None	None	female
97540	ANDRES	ERNESTO	GUERRERO	ACOSTA	None	None	None	None	male
97541	DANIELA	VILLAFañE	RINCON	None	None	None	None	None	female
97542	RICHARD	SMEINER	HERNANDEZ	RUBIO	None	None	None	None	male
97543	SERGIO	ANDRES	LOPEZ	ARAGON	None	None	None	None	male
97544	BELKI	PATRICIA	ROMERO	ESPINOSA	None	None	None	None	unknown
97545	OSNAYDER	ALBERTO	GUZMAN	GOMEZ	None	None	None	None	unknown
97546	KAREN	DAYANA	GALINDO	CARRANZA	None	None	None	None	female
97547	ANA	MARIA	OCHOA	GIRALDO	None	None	None	None	female

Fuente: Elaboración propia.

Ilustración 11. Frecuencia identificación genero primer nombre

```
[ ] 8.8 # miramos la frecuencia de la variable sexo
df3.sexo.value_counts()
```

```
female      3369
male        3009
unknown     1632
mostly_female  161
mostly_male  100
andy         5
Name: sexo, dtype: int64
```

Fuente: Elaboración propia.

De acuerdo con la identificación del género a partir del primer nombre podemos observar que se identificaron 3.369 femeninos, 3.009 masculinos, desconocidos tenemos 1.632, posibles femeninos 161 y posibles masculinos 100 nombres. Por lo cual realizamos el mismo procedimiento

para identificar el género con el segundo nombre de los compradores tabla 14. Identificación genero segundo nombre.

Tabla 14. Identificación genero segundo nombre.

```
[ ] 8.9 # determinamos el sexo guiandonos por el segundo nombre y creamos una nueva variable llamada sexo2
df3['sexo2'] = df3['segundo'].apply(str.capitalize).map(lambda x: d.get_gender(x))
df3.head(10)
```

	primer	segundo	pa	sa	a	b	c	d	sexo	sexo2
97537	ANGIEE	JULIANA	MARTINEZ	RUEDA	None	None	None	None	unknown	female
97538	NAOMI	JULIETH	CABELLO	MONTERO	None	None	None	None	female	unknown
97540	ANDRES	ERNESTO	GUERRERO	ACOSTA	None	None	None	None	male	male
97541	DANIELA	VILLAFANE	RINCON	None	None	None	None	None	female	unknown
97542	RICHARD	SMEINER	HERNANDEZ	RUBIO	None	None	None	None	male	unknown
97543	SERGIO	ANDRES	LOPEZ	ARAGON	None	None	None	None	male	male
97544	BELKI	PATRICIA	ROMERO	ESPINOSA	None	None	None	None	unknown	female
97545	OSNAYDER	ALBERTO	GUZMAN	GOMEZ	None	None	None	None	unknown	male
97546	KAREN	DAYANA	GALINDO	CARRANZA	None	None	None	None	female	female
97547	ANA	MARIA	OCHOA	GIRALDO	None	None	None	None	female	female

Fuente: Elaboración propia.

Podemos evidenciar que tenemos diferencias entre la identificación del género a partir del primer y segundo nombre por lo cual realizamos una definición definitiva del género según los resultados. Donde sí en alguno de los dos resultados de géneros identificados se encontraba como resultado desconocido, pero el otro género se identificaba se dejaba como resultado el género identificado, esto también se aplicó para los casos de que el resultado diera la opción de a lo mejor femenino o posiblemente masculino tabla. 15. Definición de género.

Tabla 15. Definición de género

```
[ ] 8.11 # Imprimimos nuestro dataset para visualizar las variables
df3.head(10)
```

	primer	segundo	pa	sa	a	b	c	d	sexo	sexo2	genero
97537	ANGIEE	JULIANA	MARTINEZ	RUEDA	None	None	None	None	unknown	female	female
97538	NAOMI	JULIETH	CABELLO	MONTERO	None	None	None	None	female	unknown	female
97540	ANDRES	ERNESTO	GUERRERO	ACOSTA	None	None	None	None	male	male	male
97541	DANIELA	VILLAFañE	RINCON	None	None	None	None	None	female	unknown	female
97542	RICHARD	SMEINER	HERNANDEZ	RUBIO	None	None	None	None	male	unknown	male
97543	SERGIO	ANDRES	LOPEZ	ARAGON	None	None	None	None	male	male	male
97544	BELKI	PATRICIA	ROMERO	ESPINOSA	None	None	None	None	unknown	female	female
97545	OSNAYDER	ALBERTO	GUZMAN	GOMEZ	None	None	None	None	unknown	male	male
97546	KAREN	DAYANA	GALINDO	CARRANZA	None	None	None	None	female	female	female
97547	ANA	MARIA	OCHOA	GIRALDO	None	None	None	None	female	female	female

Fuente: Elaboración propia.

Con lo anterior realizamos una visualización de la frecuencia definitiva de la identificación de género, evidenciando en la ilustración 12. Frecuencia identificación género, podemos observar que de nuestra base de 8.276 nombres de compradores se identificaron como femeninos 4.160, como masculinos 3.532 y desconocidos 584; Siendo el género de nuestro estudio el que más realizó compra de inmuebles el género femenino, incorporamos esta identificación de genero a nuestra `bst_filtrada`.

Ilustración 12. Frecuencia identificación género

```
[ ] 8.12 # imprimimos la frecuencia de la variable sexo definitivo
df3.genero.value_counts()
female      4160
male        3532
unknown      584
Name: genero, dtype: int64
```

Fuente: Elaboración propia.

6.4.3. Correlaciones de Variables

Para poder llevar a cabo el análisis de correlación de las diferentes variables se necesitó dummificar las variables categóricas conocidas como de texto, en variables binarias conocidas como dummy. Estas variables se usaron para el modelo de los clústeres con K-Means generando información de alto valor para la interpretación.

Al realizar el análisis de las correlaciones existentes entre las diferentes variables y la dependencia con la misma, se pudo observar que ahí una alta dependencia entre el área de los metros cuadrados de 45 a 60, perteneciendo a un tipo de producto VIP y a su vez se relaciona con el estado civil de soltero, como se evidencia en la ilustración 13

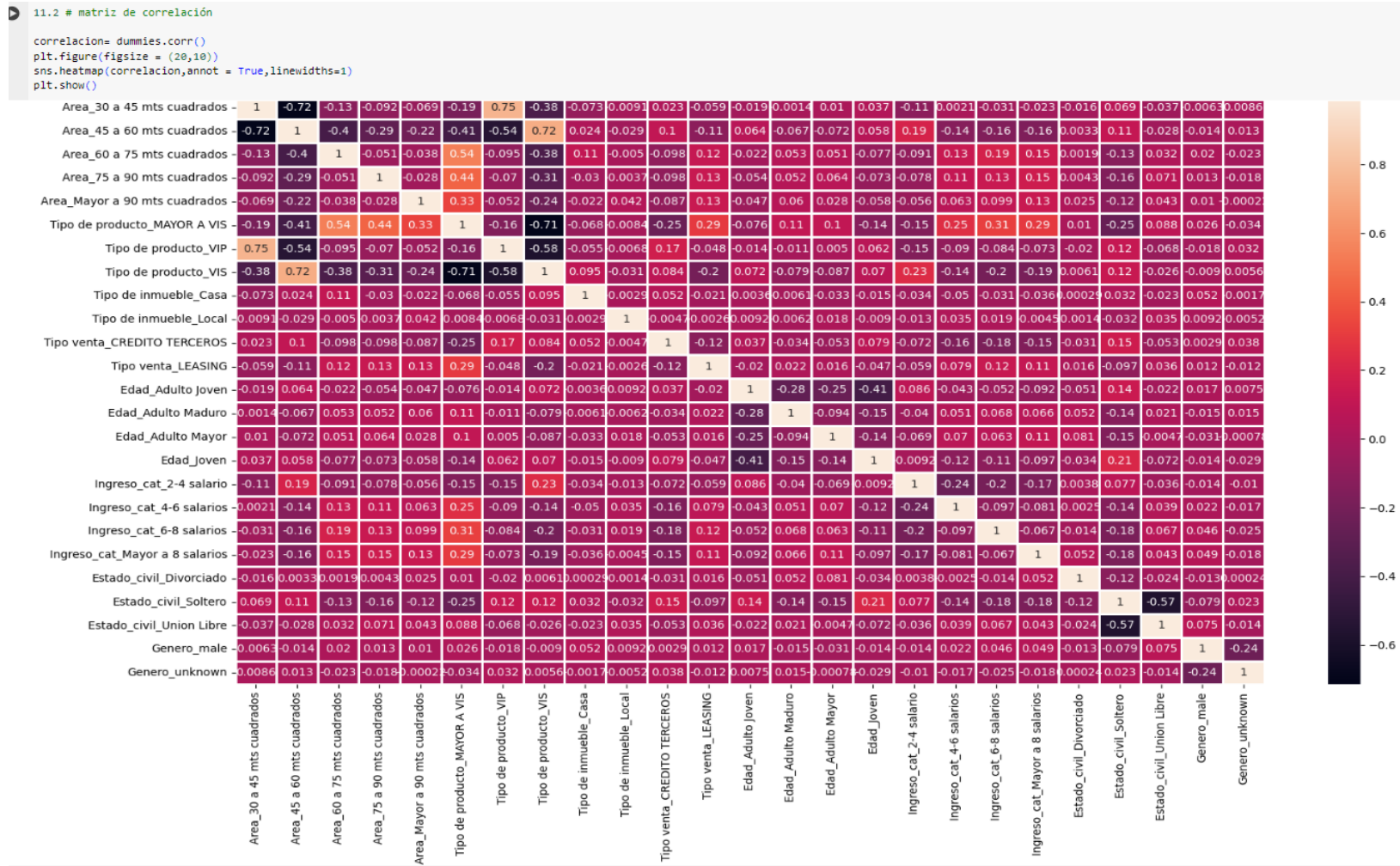
6.4.4. Implantación de K-Means

El número de clúster apropiados para un modelo de segmentación es obtenido a partir del método del codo. A través de este método, permite, hallar la cantidad de clústeres apropiados para realizar la segmentación. Este número de clústeres lo empleamos en el logaritmo de K-Means.

En un modelo no supervisado debemos analizar cuál es el número de clústeres que es óptimo a utilizar para la segmentación de los datos, esto nos permite determinar cuál es el valor ideal para K por medio del quiebre del codo “Por lo tanto, requerimos una fórmula de distancia para calcular el número eficiente de grupos para la segmentación”(Kaur & Singh, 2023).

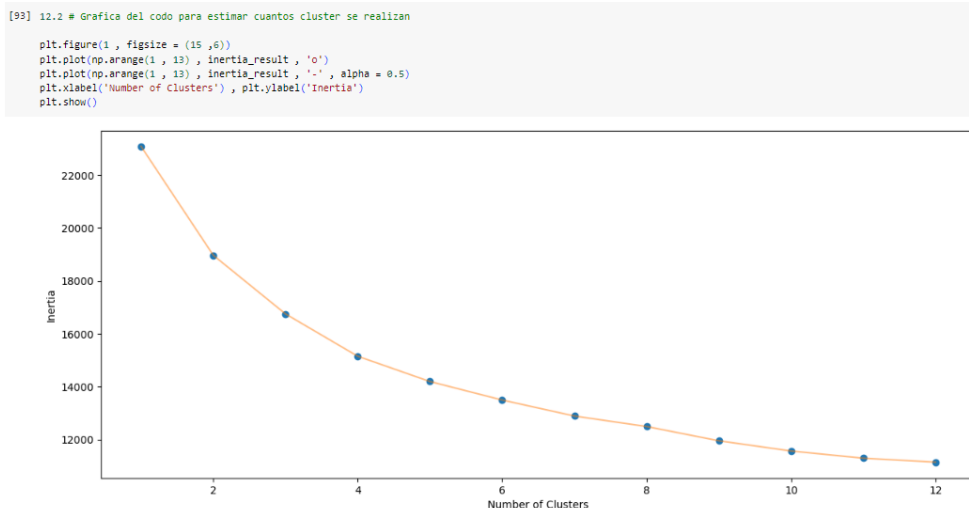
Podemos ver con la anterior ilustración 14, el resultado de cuál sería el número de clústeres ideales para nuestro modelo con respecto al quiebre del codo, se determinó que tres (3), es el número ideal con respecto a la ilustración, siendo es el número apropiado de clústeres.

Ilustración 13. Matriz de correlaciones



Fuente: Elaboración propia.

Ilustración 14. Método de codo



Fuente: Elaboración propia

Después de determinar el número ideal de clúster realizamos la indexación de los clústeres a nuestro base de datos como lo evidenciamos en la tabla 16.

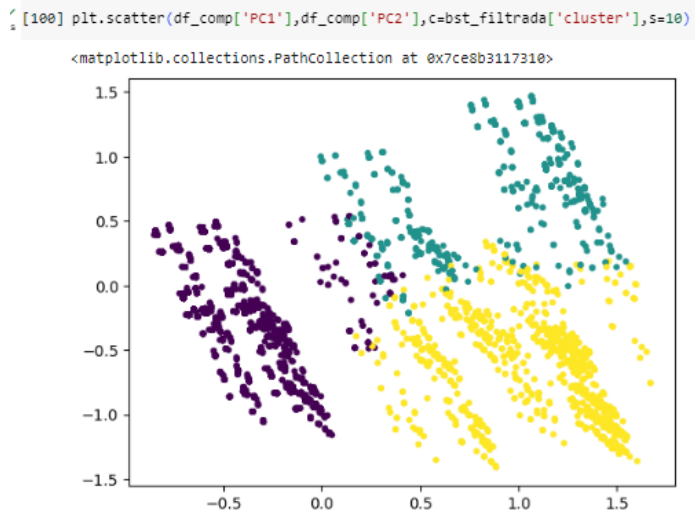
Tabla 16. Indexación de clústeres

```
[ ] bst_comp['cluster']=kmean_kpi.labels_
bst_comp.head()
```

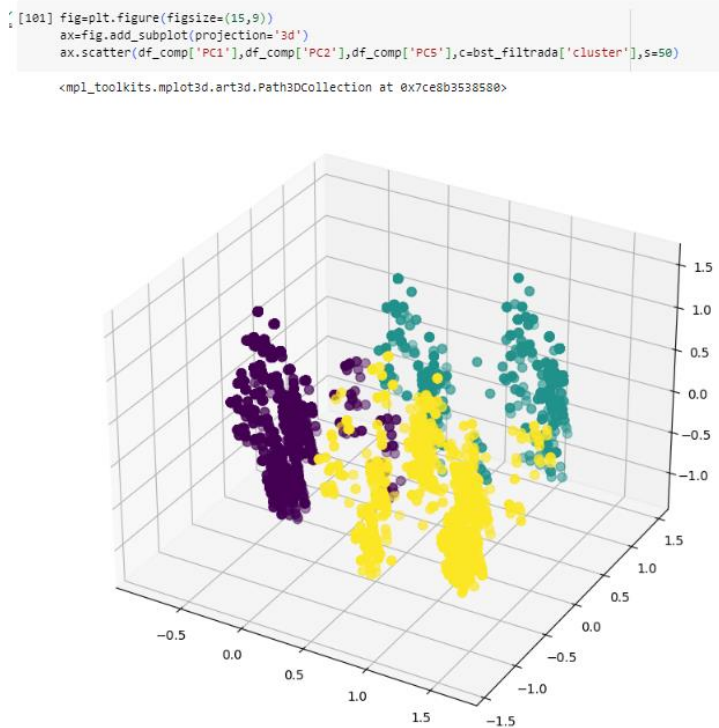
	Tipo venta	Entidad Credito	Destino	Ciudad	Area	Tipo de inmueble	Edad	cluster
97537	CREDITO	BANCO DE BOGOTA S.A.	HABITAR	Bogota D.C	45 a 60 mts cuadrados	Apartamento	Adulto Joven	0
97538	CREDITO TERCEROS	BANCO DE BOGOTA S.A.	HABITAR	Soledad	60 a 75 mts cuadrados	Casa	Joven	2
97540	CREDITO TERCEROS	BANCO CAJA SOCIAL - BCSC S.A.	HABITAR	Soacha	45 a 60 mts cuadrados	Apartamento	Adulto Joven	0
97541	CREDITO	BANCOLOMBIA S.A.	HABITAR	Villavicencio	60 a 75 mts cuadrados	Apartamento	Adulto Joven	2
97542	CREDITO	BANCO CAJA SOCIAL - BCSC S.A.	HABITAR	Ibague	30 a 45 mts cuadrados	Apartamento	Adulto Joven	2

Fuente: Elaboración propia

Para poder graficar nuestro modelo de segmentación y ver el comportamiento de los datos, necesitamos escoger dos componentes principales PCA, de la totalidad que se utilizaron en nuestro modelo donde podemos evidenciar la agrupación de los datos con etiquetas con diferentes colores observada en la ilustración 15, para un análisis más profundo podemos realizar una gráfica en 3D la cual nos permite una mejor visualización como lo observamos en la ilustración 16.

Ilustración 15. Grafica de clústeres

Fuente: Elaboración propia

Ilustración 16. Grafica de clústeres 3D

Fuente: Elaboración propia

Se optó por hacer dicha segmentación mediante el algoritmo K-Means, el cual es un enfoque popular para dividir un conjunto de datos en grupos distintos, llamados "clústeres". Estos clústeres se forman de tal manera que los puntos de datos dentro de un mismo clúster son más similares entre sí, en comparación con los puntos de datos en otros clústeres, ayudando así a que la compañía mediante su estrategia de ventas pueda identificar una segmentación adecuada buscando incrementar las ventas y llegar al cliente adecuado.

En este sentido, se pudo desarrollar con éxito un modelo de agrupación K-Means para la segmentación de clientes en la compañía donde también se pudo explorar la definición de grupos y análisis del comportamiento dentro de cada categoría, clase o grupo. Al clasificar toda la información disponible en diferentes subgrupos. Esto significa que un subgrupo o grupo tiene características que lo distinguen de otros grupos, es un conjunto de filtros similares que se utilizan para encontrar características invisibles a simple vista.

6.5.Evaluación

Durante esta etapa se llevaron a cabo diversas actividades que incluyeron comparación de diferentes modelos, para determinar la idoneidad y el rendimiento de los modelos creados durante este proyecto, ayudando a garantizar que los modelos sean útiles y efectivos para abordar el problema empresarial de investigación.

Se realizaron las siguientes tres (3) clasificaciones por tipo de evento, donde nuestra primera clasificación es por el evento de compra, esta clasificación incluye las variables de tipo de venta, entidad de crédito, destino, ciudad, área del inmueble, tipo de inmueble y la edad del comprador como lo muestra la tabla 17.

Tabla 17. Tipo de evento compra

Compra

```
[ ] bst_comp['cluster']=kmean_kpi.labels_
bst_comp.head()
```

	Tipo venta	Entidad Credito	Destino	Ciudad	Area	Tipo de inmueble	Edad	cluster
97537	CREDITO	BANCO DE BOGOTA S.A.	HABITAR	Bogota D.C	45 a 60 mts cuadrados	Apartamento	Adulto Joven	0
97538	CREDITO TERCEROS	BANCO DE BOGOTA S.A.	HABITAR	Soledad	60 a 75 mts cuadrados	Casa	Joven	2
97540	CREDITO TERCEROS	BANCO CAJA SOCIAL - BCSC S.A.	HABITAR	Soacha	45 a 60 mts cuadrados	Apartamento	Adulto Joven	0
97541	CREDITO	BANCOLOMBIA S.A.	HABITAR	Villavicencio	60 a 75 mts cuadrados	Apartamento	Adulto Joven	2
97542	CREDITO	BANCO CAJA SOCIAL - BCSC S.A.	HABITAR	Ibague	30 a 45 mts cuadrados	Apartamento	Adulto Joven	2

Fuente: Elaboración propia

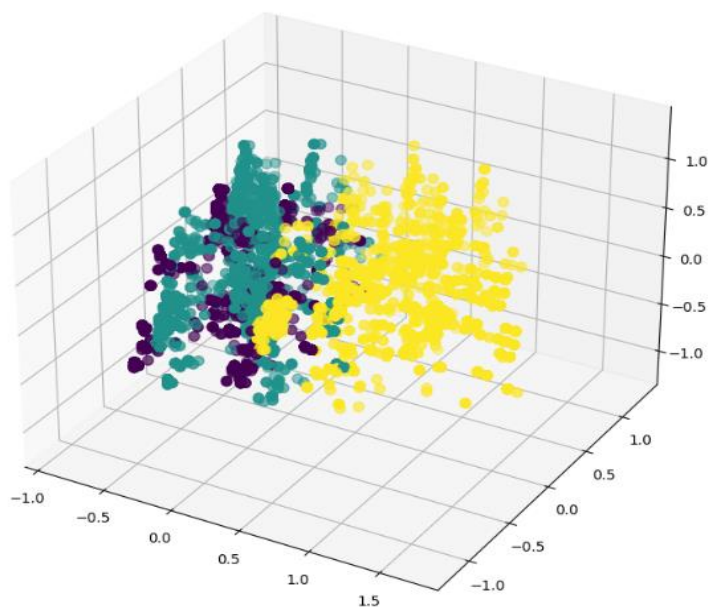
En este tipo de evento compra en la ilustración 17, podemos visualizar los clústeres que se formaron con etiquetas de colores, la distribución de los puntos de datos en el espacio tridimensional se puede identificar los patrones de agrupamiento.

Ilustración 17. Clúster evento compra

```
[ ] bst_comp.to_csv('cluster_comp.csv')

[ ] fig=plt.figure(figsize=(15,9))
ax=fig.add_subplot(projection='3d')
ax.scatter(df_comp['PC1'],df_comp['PC2'],df_comp['PC7'],c=bst_comp['cluster'],s=50)

<matplotlib.pyplot.Path3DCollection at 0x7ce8b2bcdcc0>
```



Fuente: Elaboración propia

En la segunda clasificación es por el evento de persona, esta clasificación incluye las variables de género, Ingresos, fuente, edad, estado civil, ocupación y tipo de producto como lo muestra la tabla 18

Tabla 18. Clúster evento persona

Persona

```
bst_per['cluster']=kmean_kpi.labels_
bst_per.head()
```

	Genero	Ingreso_cat	Fuente	Edad	Estado_civil	Ocupación	Tipo de producto	cluster
97537	female	2-4 salario	LEADS	Adulto Joven	Soltero	Empleado	VIS	2
97538	female	2-4 salario	LEADS	Joven	Soltero	Empleado	VIS	1
97540	male	0-2 salarios	PAGINA_WEB_AMARILO	Adulto Joven	Soltero	Empleado	VIS	2
97541	female	4-6 salarios	REFERIDO	Adulto Joven	Soltero	Empleado	MAYOR A VIS	0
97542	male	0-2 salarios	VALLA_SALA_DE_VENTAS	Adulto Joven	Soltero	Empleado	VIP	2

Fuente: Elaboración propia

En este tipo de evento persona en la ilustración 18, podemos visualizar los clústeres que se formaron con etiquetas de colores, la distribución de los puntos de datos en el espacio tridimensional se puede identificar los patrones de agrupamiento, donde se tuvieron en cuenta las variables más representativas con características de las personas compradoras de inmuebles. En la ilustración observamos que la distribución de los grupos es más homogénea y compacta, formando clústeres más esféricos al tener concentraciones de puntos cercanos entre sí.

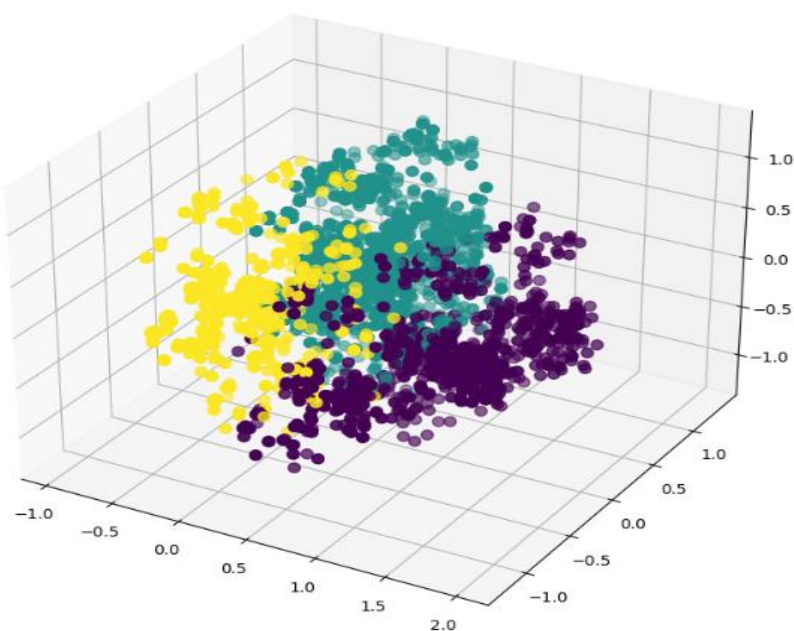
Para nuestra tercera clasificación se realizó por el evento inmueble, esta clasificación incluye las variables de edad del comprador, ciudad, área, tipo de producto, tipo de inmueble y estado civil como lo muestra la tabla 19.

Ilustración 18. Clúster evento persona

```
[ ] bst_per.to_csv('cluster_perso.csv')

[ ] fig=plt.figure(figsize=(15,9))
ax=fig.add_subplot(projection='3d')
ax.scatter(df_comp['PC1'],df_comp['PC2'],df_comp['PC6'],c=bst_per['cluster'],s=50)

<mpl_toolkits.mplot3d.art3d.Path3DCollection at 0x7ce8b34ea110>
```



Fuente: Elaboración propia

Tabla 19. Clúster evento inmueble

Inmueble

```
[ ] bst_inm['cluster']=kmean_kpi.labels_
bst_inm.head()
```

	Edad	Ciudad	Area	Tipo de producto	Tipo de inmueble	Estado_civil	cluster
97537	Adulto Joven	Bogota D.C	45 a 60 mts cuadrados	VIS	Apartamento	Soltero	1
97538	Joven	Soledad	60 a 75 mts cuadrados	VIS	Casa	Soltero	1
97540	Adulto Joven	Soacha	45 a 60 mts cuadrados	VIS	Apartamento	Soltero	1
97541	Adulto Joven	Villavicencio	60 a 75 mts cuadrados	MAYOR A VIS	Apartamento	Soltero	2
97542	Adulto Joven	Ibague	30 a 45 mts cuadrados	VIP	Apartamento	Soltero	0

Fuente: Elaboración propia

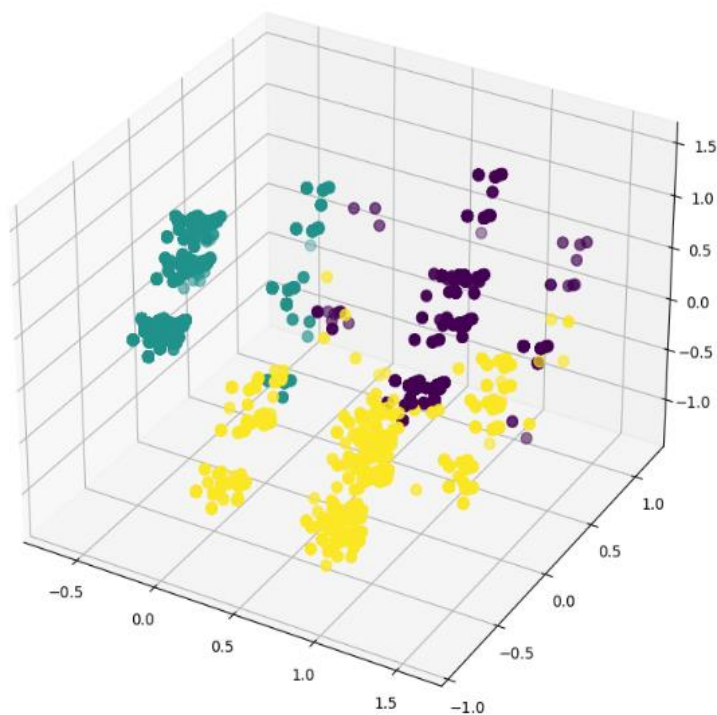
En este tipo de evento inmueble en la ilustración 19, podemos visualizar los clústeres que se formaron con etiquetas de colores, la distribución de los puntos de datos en el espacio tridimensional se puede identificar los patrones de agrupamiento, donde se tuvieron en cuenta las variables más representativas con características de los inmuebles. En la ilustración observamos que la distribución de los grupos es más separada entre sí, formando figuras irregulares al tener concentraciones de puntos distantes entre sí.

Ilustración 19. Clúster evento inmueble

```
[ ] bst_inm.to_csv('cluster_inmu.csv')

[ ] fig=plt.figure(figsize=(15,9))
    ax=fig.add_subplot(projection='3d')
    ax.scatter(df_comp['PC1'],df_comp['PC4'],df_comp['PC5'],c=bst_inm['cluster'],s=50)

<mpl_toolkits.mplot3d.art3d.Path3DCollection at 0x7ce8b30019c0>
```



Fuente: Elaboración propia

Realizamos una consulta para medir la relación lineal entre las variables cuantitativas Área, Ingresos, Edad y Vr Neto en correlación con el tipo de producto que ofrece la compañía el cual nos arrojó la ilustración 20, donde se observa la forma de agrupación, comportamiento y distribución de los cuatro tipos de producto existentes.

Ilustración 20. Correlación entre variables



Fuente: Elaboración propia

En esta Ilustración 21 se puede evidenciar que la prevalencia del adulto joven es la que mejor busca adquirir vivienda, sin embargo, el rango de adulto también posee un incremento importante en cuanto a la adquisición de viviendas, asimismo, se puede apreciar que la población adulto mayor no enmarca un segmento adecuado para dicho proceso.

Tabla 20. Distribución de edad

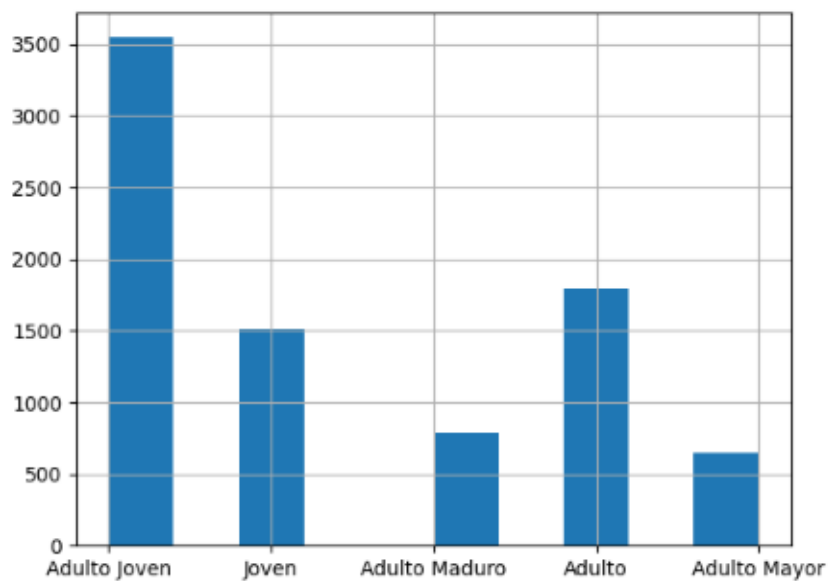
Tipo de persona	Rango Edad	Frecuencia
Adulto Joven	Entre 26 y 35 años	3.544
Adulto	Entre 36 y 45 años	1.794
Joven	Entre 18 y 25 años	1.508
Adulto Maduro	Entre 46 y 55 años	786
Adulto Mayor	Mas de 56	644
Total		8.276

Fuente: Elaboración propia

Ilustración 21. Grafica de edades

```
[71] 10.5 # Histograma visualización Univariada de la información
bst_filtrada['Edad'].hist()
```

<Axes: >



Fuente: Elaboración propia

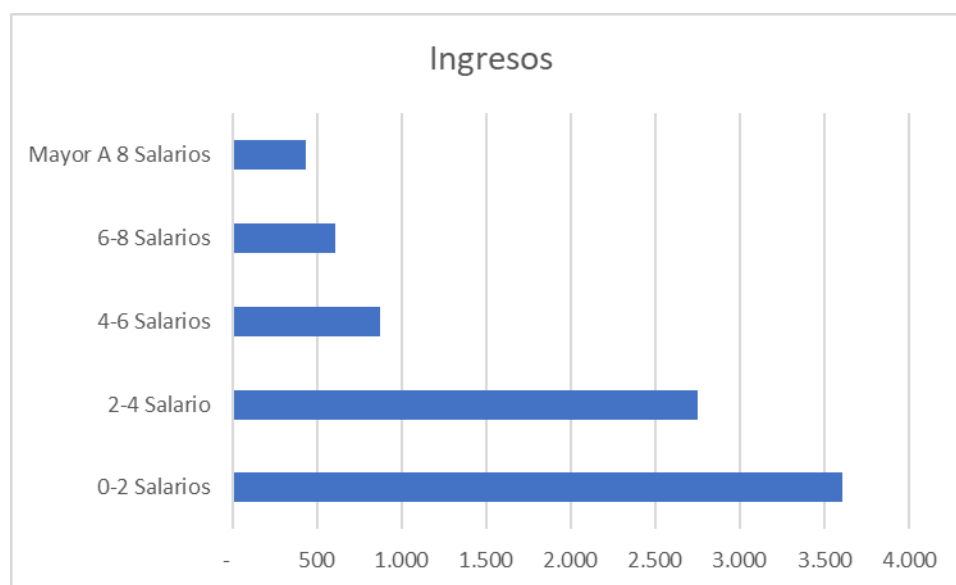
En la Ilustración 22 se observa el segmento de ingresos que arrojaron los datos de estudio, el cual las personas que ganan hasta 2 salarios mínimos son las que más adquieren viviendas, en relación con los que ganan entre 2 y 4 salarios. Asimismo, las personas que tienen un salario mayor a 8 enmarcan pocas necesidades de adquirir una vivienda en cuanto a que tienen otras proyecciones, por eso es importante describir mediante este algoritmo identificar el tipo de ingreso que no adquiere vivienda y asimismo lograr abordar ese nicho

Tabla 21. Rango de ingresos

Ingresos	Frecuencia
0-2 Salarios	3.605
2-4 Salario	2.749
4-6 Salarios	876
6-8 Salarios	611
Mayor A 8 Salarios	435
Total	8.276

Fuente: Elaboración propia

Ilustración 22. Grafica de ingresos



Fuente: Elaboración propia

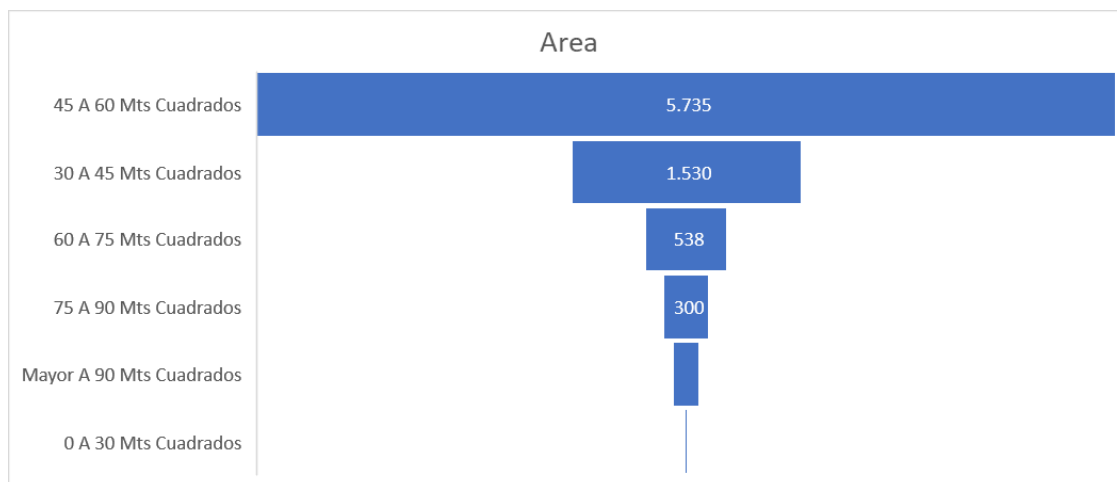
Se observa que el área más solicitada por parte de los clientes es la de 45 a 60 mts cuadrados, debido a que cumple con las expectativas en cuanto a su distribución. Teniendo en cuenta el análisis, se busca llevar a cabo una estrategia adecuada para que las viviendas menos vendidas puedan tener un mayor impulso en las ventas.

Tabla 22. Distribución de área

Area	Frecuencia
45 A 60 Mts Cuadrados	5.735
30 A 45 Mts Cuadrados	1.530
60 A 75 Mts Cuadrados	538
75 A 90 Mts Cuadrados	300
Mayor A 90 Mts Cuadrados	171
0 A 30 Mts Cuadrados	2
Total	8.276

Fuente: Elaboración propia

Ilustración 23. Grafica área de inmuebles



Fuente: Elaboración propia

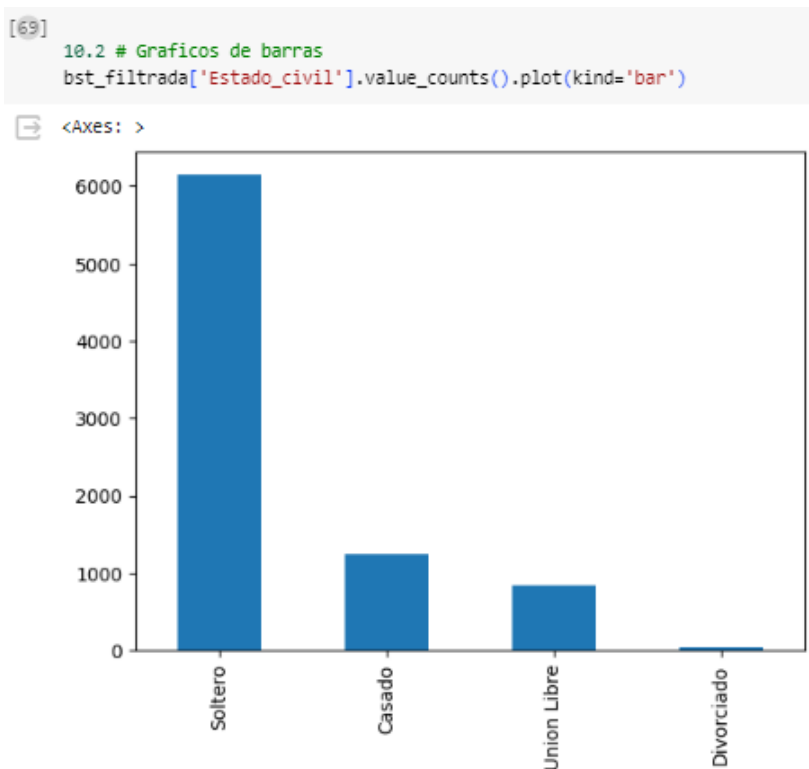
En la Ilustración 24 se puede evidenciar que las personas solteras son las que más tienen intención de adquirir una vivienda, sin embargo, las personas casadas enmarcan también una frecuencia importante a la hora de adquirir un inmueble.

Tabla 23. Estado civil

Estado Civil	Frecuencia
Soltero	6.141
Casado	1.243
Union Libre	850
Divorciado	42
Total	8.276

Fuente: Elaboración propia

Ilustración 24. Grafica estado civil



Fuente: Elaboración propia

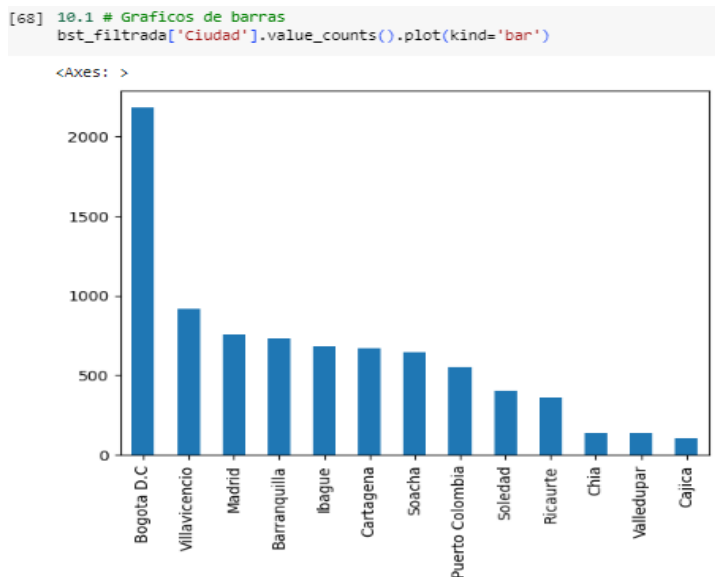
Observamos que la ciudad con más venta de viviendas es Bogotá, seguida de Villavicencio y Madrid en Cundinamarca, estas ciudades son focos estratégicos para el crecimiento de la compañía. Enmarcando en análisis de datos se puede llevar a cabo estrategias de mercado para la ciudad con menos adquisición de viviendas, logrando tener una mejor utilidad para la compañía.

Tabla 24. Distribución por ciudad

Ciudad	Frecuencia
Bogota D.C	2.177
Villavicencio	917
Madrid	757
Barranquilla	731
Ibague	681
Cartagena	668
Soacha	644
Puerto Colombia	550
Soledad	405
Ricaurte	360
Chia	140
Valledupar	138
Cajica	10
Total	8.178

Fuente: Elaboración propia

Ilustración 25. Grafica por ciudad



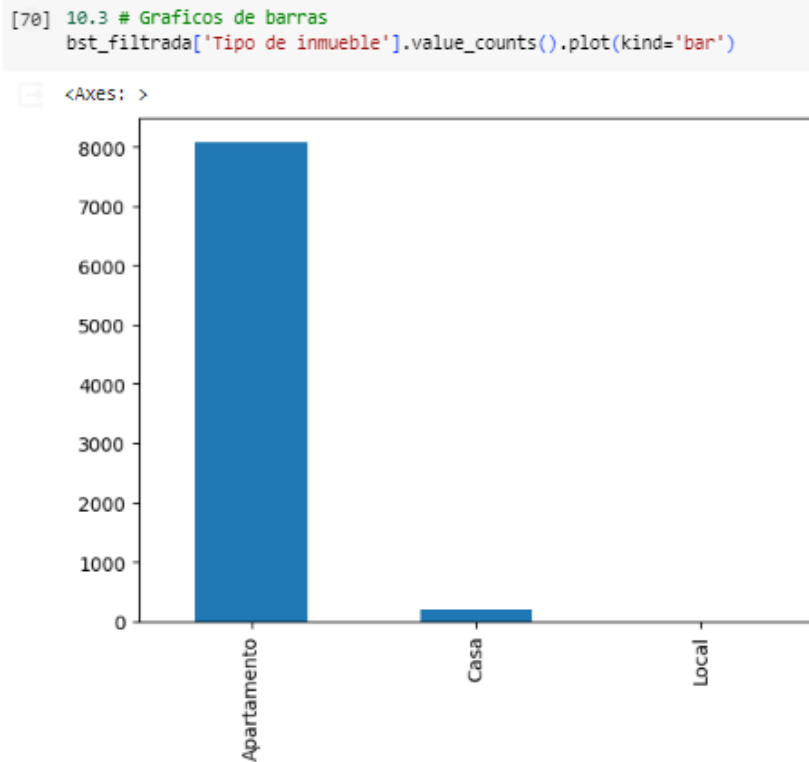
Fuente: Elaboración propia

Se observa en la ilustración 26, que el apartamento tiene gran relevancia en el proceso de compra del cliente, en relación con las casas, ya que en la actualidad las personas optan por adquirir un apartamento, el cual es cómodo y seguro, siendo impulsado por los programas de beneficios en la adquisición de vivienda que otorga el estado.

Tabla 25. Tipo de inmueble

Inmueble	Frecuencia
Apartamento	8.081
Casa	192
Local	3
Total	8.276

Fuente: Elaboración propia

Ilustración 26. Grafica del tipo de inmueble

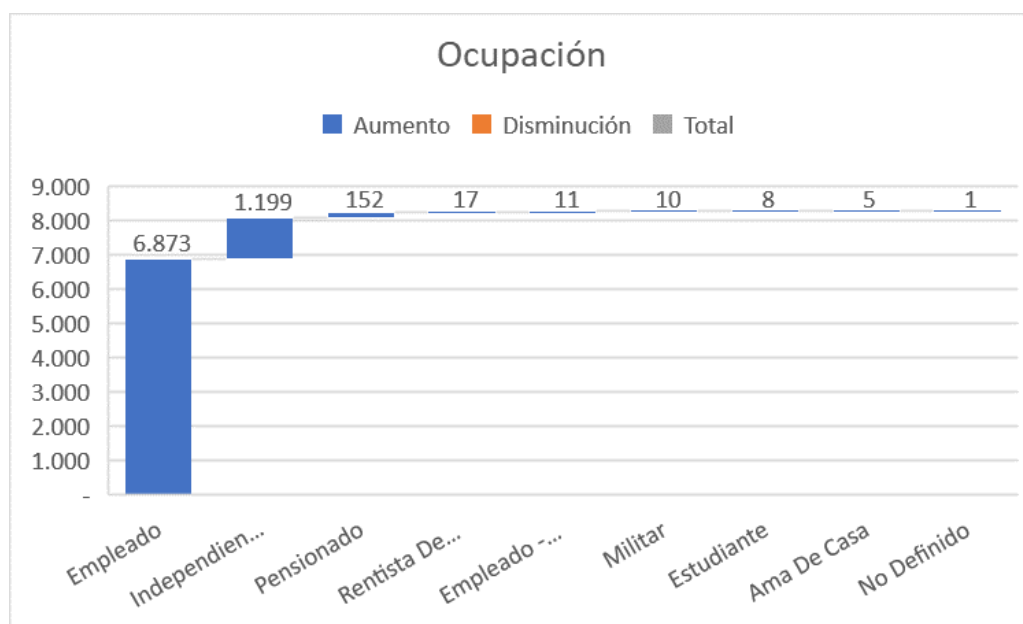
Fuente: Elaboración propia

En la Ilustración 27 se puede observar que usualmente el empleado es quien adquiere la vivienda, debido a que su sueldo y tipo de contrato es seguro para la constructora y el banco al momento de abordar las cuotas

Tabla 26. Distribución por Ocupación

Ocupación	Frecuencia
Empleado	6.873
Independiente	1.199
Pensionado	152
Rentista De Capital	17
Empleado - Socio	11
Militar	10
Estudiante	8
Ama De Casa	5
No Definido	1
Total	8.276

Fuente: Elaboración propia

Ilustración 27. Grafica Ocupación

Fuente: Elaboración propia

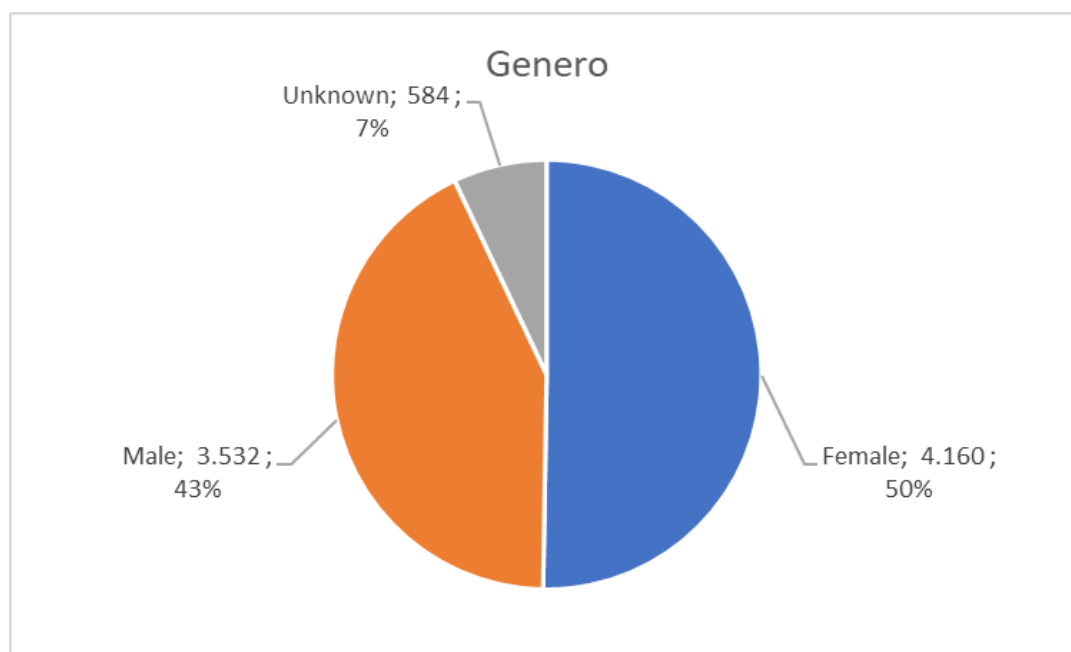
En esta ilustración observamos que el género femenino es la principal adquisidora de inmuebles ya que representa el 50% con un total de 4.160 compradores, dentro de nuestro análisis el género masculino representa el 43% con 3.532 compradores y sin identificar el género un 7% que equivalen a 584 compradores.

Tabla 27. Distribución de genero

Genero	Frecuencia
Female	4.160
Male	3.532
Unknown	584
Total	8.276

Fuente: Elaboración propia

Ilustración 28. Grafica genero



Fuente: Elaboración propia

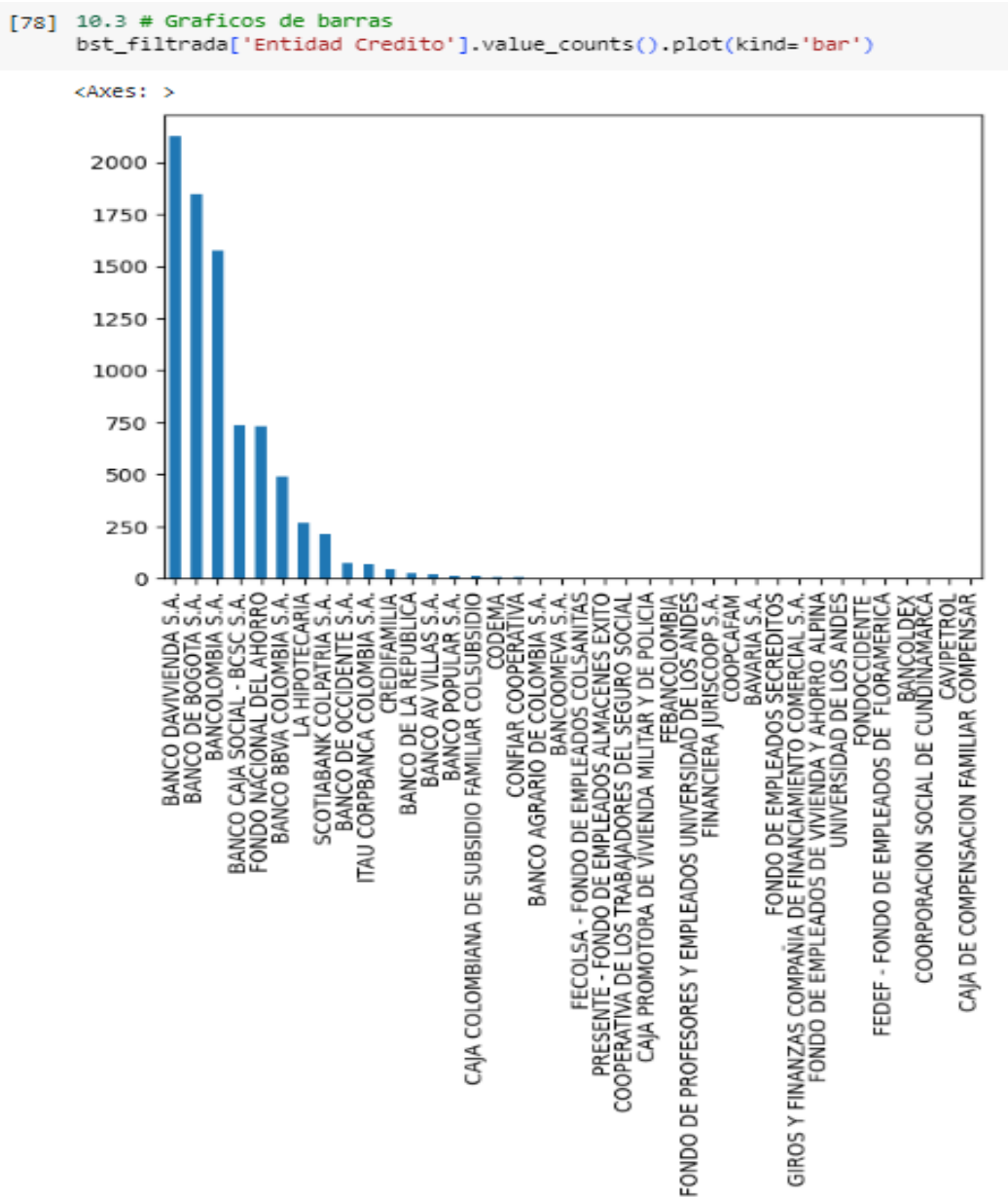
En la ilustración 29, observamos que la entidad crediticia que más otorga créditos para la adquisición de inmuebles es el banco Davivienda, seguido del banco de Bogotá y Bancolombia, siendo estas tres entidades financieras las más representativas, lo cual nos permite poder crear alianzas estratégicas entre la compañía y las entidades financieras para impulsar la adquisición de viviendas

Tabla 28. Distribución entidad de crédito

Entidad Credito	Frecuencia
Banco Davivienda S.A.	2.123
Banco De Bogota S.A.	1.845
Bancolombia S.A.	1.574
Banco Caja Social - Bcsc S.A.	736
Fondo Nacional Del Ahorro	729
Banco Bbva Colombia S.A.	488
La Hipotecaria	266
Scotiabank Colpatría S.A.	211
Banco De Occidente S.A.	73
Itau Corpbanca Colombia S.A.	67
Credifamilia	42
Banco De La Republica	23
Banco Av Villas S.A.	22
Caja Colombiana De Subsidio Familiar Colsubsidio	16
Banco Popular S.A.	16
Codema	7
Confiar Cooperativa	6
Banco Agrario De Colombia S.A.	4
Bancoomeva S.A.	4
Caja Promotora De Vivienda Militar Y De Policia	2
Cooperativa De Los Trabajadores Del Seguro Social	2
Presente - Fondo De Empleados Almacenes Exito	2
Febancolombia	2
Fecolsa - Fondo De Empleados Colsanitas	2
Fondo De Empleados Secreditos	1
Giros Y Finanzas Compañía De Financiamiento Comercial S.A.	1
Fondoccidente	1
Fondo De Profesores Y Empleados Universidad De Los Andes	1
Coopcafam	1
Fondo De Empleados De Vivienda Y Ahorro Alpina	1
Financiera Juriscoop S.A.	1
Fedef - Fondo De Empleados De Floramerica	1
Coorporacion Social De Cundinamarca	1
Cavipetrol	1
Caja De Compensacion Familiar Compensar	1
Bavaria S.A.	1
Bancoldex	1
Universidad De Los Andes	1
Total	8.276

Fuente: Elaboración propia

Ilustración 29. Grafica entidad de crédito



Fuente: Elaboración propia

En esta ilustración 30, observamos que la mayoría de nuestros compradores de vivienda, adquirieron sus inmuebles por la fuente de referidos, lo cual es un canal para realizarle seguimiento desde marketing y continuar a través de nuestros clientes siendo referido

Tabla 29. Distribución de la fuente

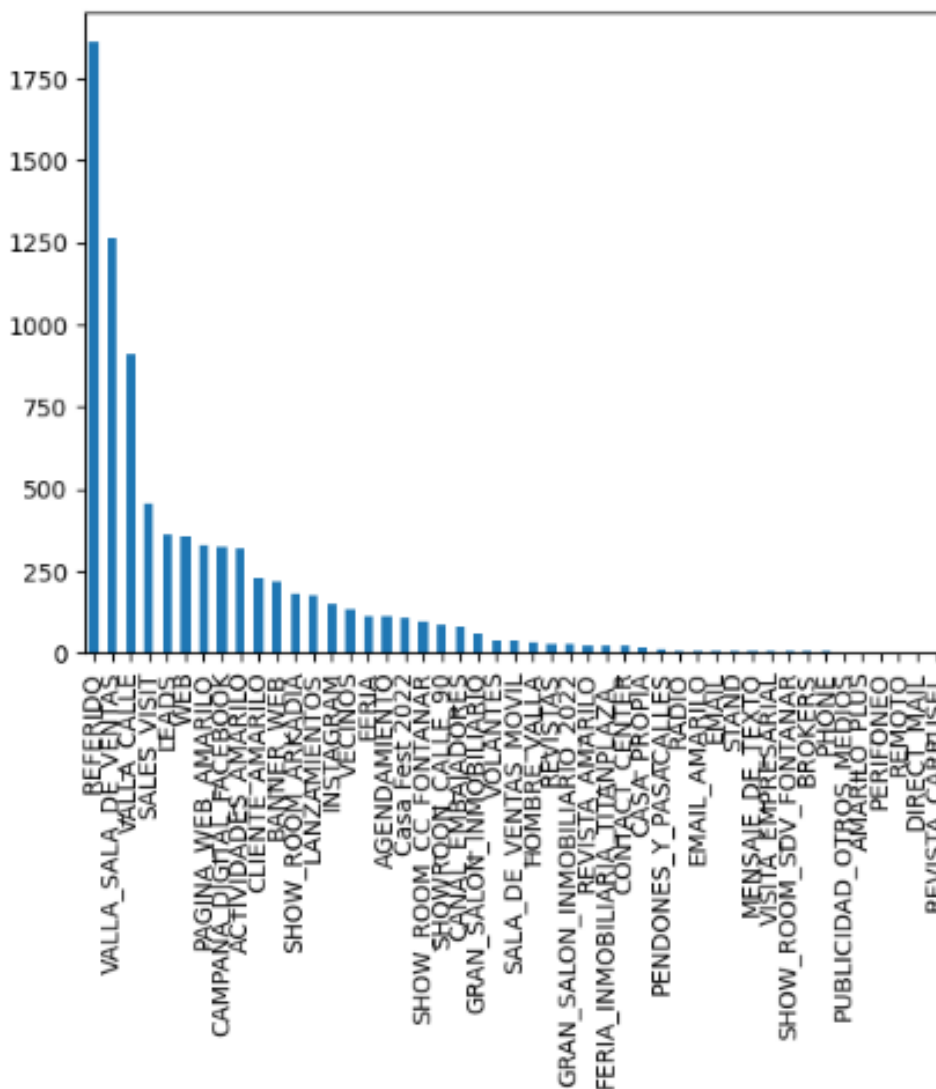
Fuente	Frecuencia
Referido	1.858
Valla_Sala_De_Ventas	1.264
Valla_Calle	911
Sales_Visit	454
Leads	363
Web	356
Pagina_Web_Amarilo	327
Campaña_Digital_Facebook	323
Actividades_Amarilo	318
Cliente_Amarilo	230
Banner_Web	216
Show_Room_Arkadia	184
Lanzamientos	177
Instagram	152
Vecinos	133
Feria	112
Agendamiento	112
Casa Fest 2022	109
Show_Room_Cc_Fontanar	97
Showroom_Calle_90	87
Canal_Embajadores	82
Gran_Salon_Inmobiliario	62
Volantes	41
Sala_De_Ventas_Movil	39
Hombre_Valla	33
Revistas	30
Gran_Salon_Inmobiliario_2022	29
Revista_Amarilo	25
Feria_Inmobiliaria_Titanplaza_	23
Contact_Center	23
Casa_Propia	16
Pendones_Y_Pasacalles	12
Radio	10
Email_Amarilo	9
Stand	8
Email	8
Mensaje_De_Texto	7
Visita_Empresarial	7
Show_Room_Sdv_Fontanar	6
Phone	5
Brokers	5
Publicidad_Otros_Medios	4
Amarilo_Plus	3
Remoto	2
Perifoneo	2
Revista_Carrusel	1
Direct_Mail	1
Total	8.276

Fuente: Elaboración propia

Ilustración 30. Grafica de fuente

```
[77] 10.3 # Graficos de barras
bst_filtrada['Fuente'].value_counts().plot(kind='bar')
```

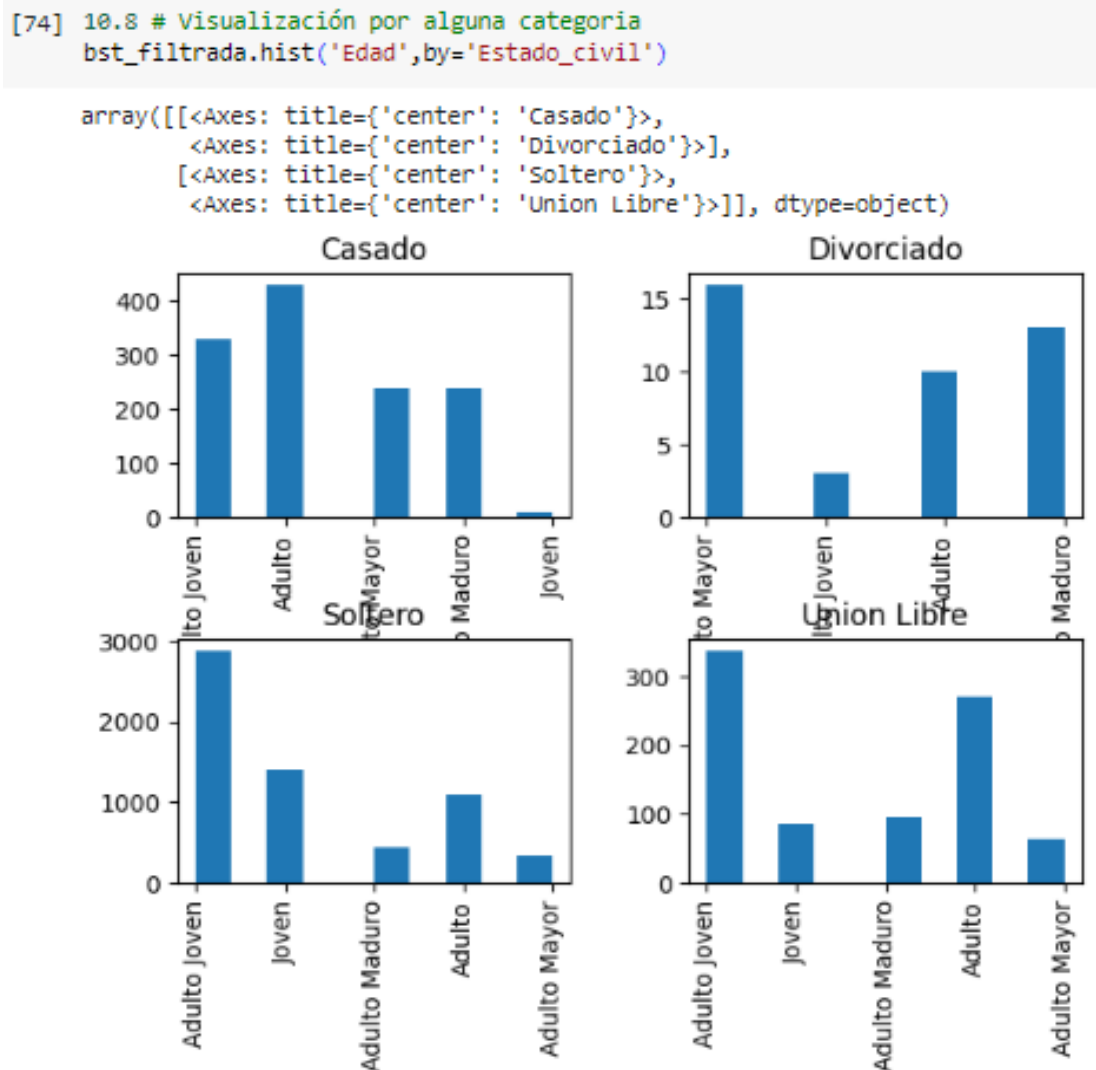
<Axes: >



Fuente: Elaboración propia

En la ilustración 31, realizamos una visualización grafica de las variables edad y estado civil, permitiéndonos analizar sobre el estado civil que tienen nuestros compradores de vivienda y en que rango de edad se encontraban en el momento de la adquisición del inmueble.

Ilustración 31. Grafica relación edad y estado civil



Fuente: Elaboración propia

6.6.Despliegue

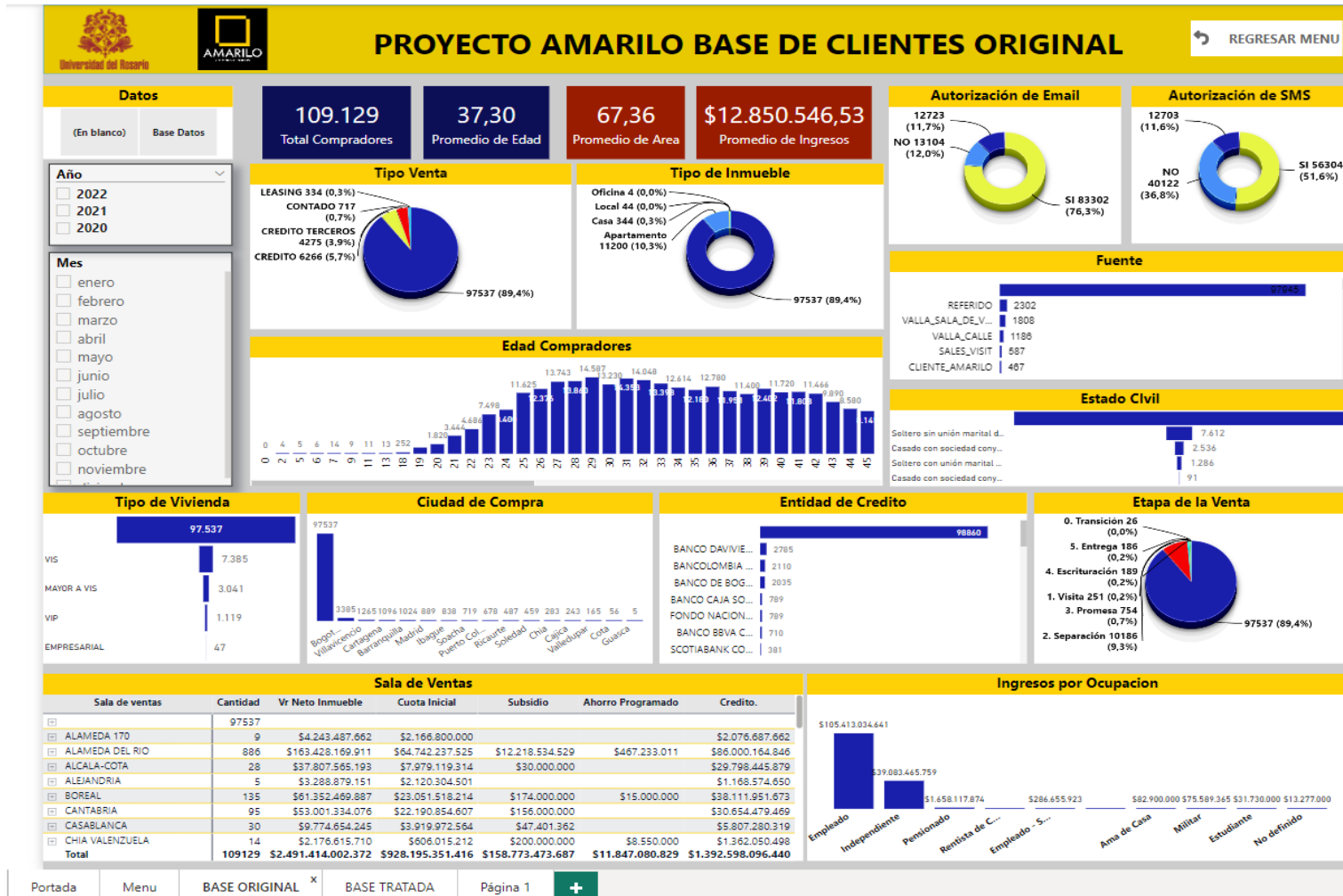
En la sexta fase llamada etapa de despliegue teniendo en cuenta la metodología de Crisp-MD, se creó un tablero de visualización Dashboard en el programa de Power Bi para que los datos de las características de las variables sean más fáciles de entender tanto para el área de marketing como para los grandes directivos de la compañía, siendo una herramienta clave para la toma de decisiones.

El propósito del tablero de visualización es proporcionar la información en tiempo real, siendo visualmente atractiva sobre el estado actual de las características de los clientes que están interesados en adquirir un inmueble y oportunidades comerciales, a través de este Dashboard podemos ver los grupos que comparten en común intereses y patrones similares a la hora de adquirir un inmueble, siendo útil para el área de marketing generar campañas segmentadas de la compañía.

El Dashboard se realizó en dos pestañas del informe, en la primera pestaña es un tablero de visualización con todos los datos sin realizarles limpieza o tratamiento, donde podemos observar la cantidad de datos que se encuentran nulos y que afectan nuestro informe como se observa en la ilustración 32.

En la segunda pestaña del informe encontramos un tablero de visualización con los datos tratados, higienización de datos y eliminación de outliers, donde podemos observar que los datos nos muestran información más precisa sobre las características y gustos de las personas interesadas en la adquisición de un inmueble, se evidencia en la ilustración 33.

Ilustración 32. Dashboard sin tratamiento de datos



Fuente: Elaboración propia

Ilustración 33. Dashboard con tratamiento de datos



Fuente: Elaboración propia

7. Descripción de las alternativas, estrategias y/o acciones que se toman en el análisis de la solución a la problemática

El mercado de la construcción en Colombia es muy diverso y en constante cambio, ya que se ve afectado por factores externos que afectan su disponibilidad y demanda. Factores políticos, sociales, económicos, ambientales y otros interfieren en el desarrollo de este proyecto, buenas cifras en estas áreas indican buenas actividades de construcción, debido a que, es uno de los movimientos reformistas más poderosos de la economía mundial.

Reconociendo el poder de la industria de la construcción en la economía, los gobiernos pertinentes han introducido diversos incentivos mediante diversos planes para acelerar la construcción y venta de inmuebles, como préstamos y asistencia en el desarrollo de planes regionales, planes, tasas hipotecarias más bajas. Por ende, las empresas sólo pueden crear rápidamente métodos y entregar los productos que los clientes desean a través del diseño, la responsabilidad, la calidad, la gestión del cliente y la estrategia (Arango & Otero 2018).

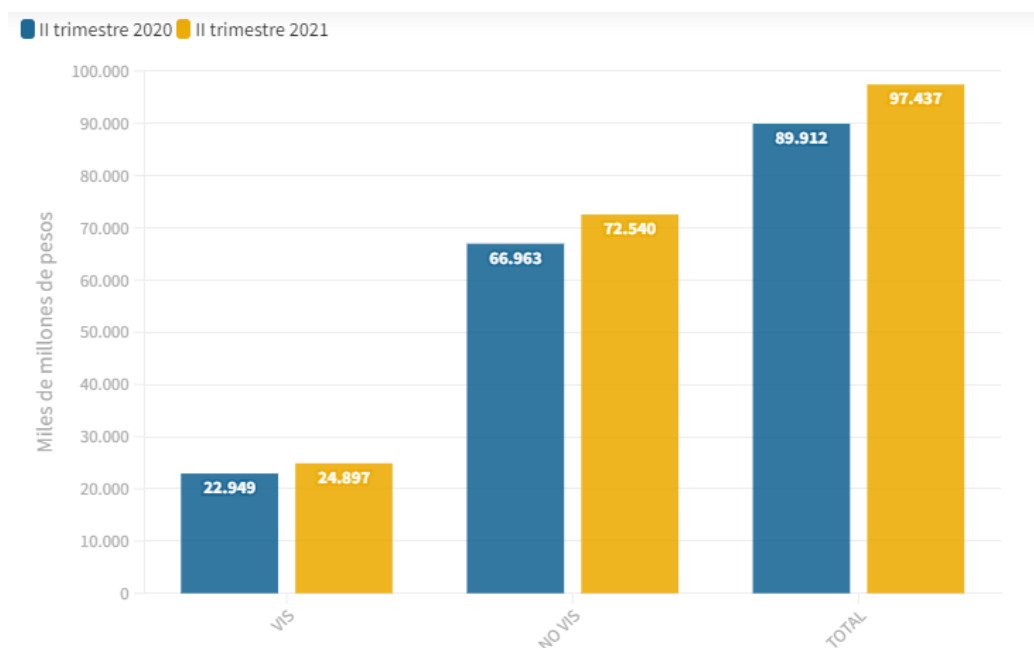
Según Crollo & Sanenz (2018), debido a la buena actividad económica y constructiva en 2019, el interés por encontrar vivienda sigue aumentando. Según CAMACOL se espera que la economía crezca un 2,6%, superior a la tasa de crecimiento del 1,7% de 2017. La recuperación económica se está acelerando, impulsada por un retorno al equilibrio de los ingresos de los hogares.

Es por esto la importancia para la compañía, ya que la finalidad de determinar el problema que afecta el posicionamiento y ubicación en el mercado, se buscan que las estrategias de marketing adecuadas puedan aumentar sus ventas, ayudarle a ganar un nicho aumentando la popularidad de su empresa en el mercado.

En el segundo trimestre de 2023 (de abril a junio), el PIB en precios reales aumentó un 0,3% en comparación con el primer trimestre de 2022. Analizando los resultados del valor añadido

de las principales industrias, el valor añadido de la industria manufacturera cayó un 3,7% según el ciclo anual, debido principalmente a la disminución del valor añadido de los proyectos públicos (-17,9%); el valor añadido de los servicios privados cayó un 3,2% y el valor añadido de los edificios aumentó un 3,1% (DANE 2023).

Ilustración 34. Saldo de capital total de la cartera hipotecaria



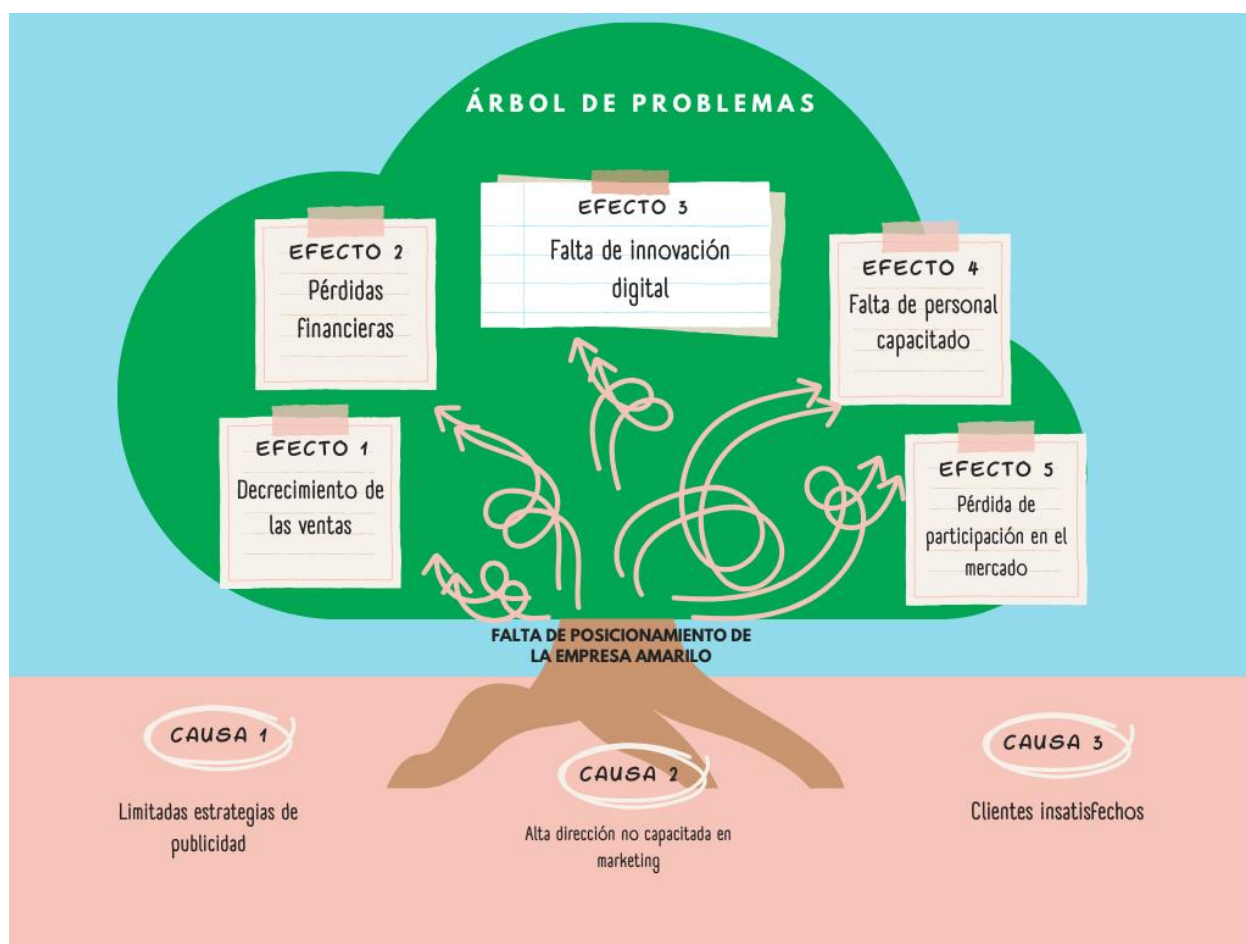
Fuente: DANE

En Colombia, todas las empresas con diferentes nombres corporativos, especialmente los altos directivos, esperan aumentar las ventas mediante el desarrollo activo de diferentes grupos en beneficio de toda la situación gerencial, especialmente de los propietarios y gerentes, mediante la capacidad para gestionar procesos para segmentos de productos y servicios de los mercados en los que participa la empresa.

Por lo tanto, es muy importante que la empresa en estudio utilice una variedad de estrategias de marketing para lograr un buen posicionamiento teniendo en cuenta la estrategia de

siete partes: participación de mercado, competencia, clientes, contabilidad, cantidad y producción, estructura de información y proveedores, con el objetivo de que el equipo de la compañía comprenda e identifique problemas que limitan la distribución, desarrollo mayorista de bienes y desarrolle estrategias de marketing adecuadas, vendiendo lo básico sin perder producto, habilidades, personas y dinero.

Ilustración 35. Árbol de problemas



Fuente: Elaboración propia (Canvas)

7.1. Análisis crítico

El problema encontrado por la compañía es la falta de seguimiento de ventas, lo que afecta el crecimiento de la empresa debido a que le resulta difícil mantenerse en el mercado, se puede analizar que una de las razones del problema es la falta de métodos de marketing, ya que, los métodos de marketing no se utilizan correctamente tal y, como lo han demostrado las investigaciones en el campo, esto resulta en una disminución en las ventas.

El uso de estrategias de marketing aumentará las ventas, pero para ello primero se debe identificar el segmento de mercado al que se dirige el producto; luego, a partir del análisis se pueden formular estrategias en función de su edad, género, ingresos, etc.

7.2. Estrategias

Actualmente, la compañía cuenta con un adecuado músculo financiero a nivel nacional con una experiencia óptima en el sector, asimismo, la ciudad se encuentra en un momento de crecimiento urbano y expansión hacia las áreas de influencia, convirtiéndose en un momento oportuno para que la organización impacte en el mercado de la construcción.

Teniendo en cuenta las necesidades del mercado por el cambio generacional con el fin de no diversificar el portafolio y continuar con una oferta concentrada únicamente en el sector residencial, podría resultar nefasto para la proyección de la empresa a futuro, perdiendo competitividad y participación del mercado.

7.2.1. Estrategias financieras

- Diversificar el portafolio con una atractiva oferta que permita ganar mayor participación en el mercado
- Desarrollar alianzas público-privadas enfocadas en cambiar el déficit de vivienda en la ciudad

- Trabajar en estrecha colaboración con Casa Propia Colombia para crear un canal de ventas directas para los colombianos que viven en el exterior.
- Aprovechamiento del programa de mi Casa Ya subsidio a la tasa de interés para potencializar las ventas de los inmuebles superiores a VIS, el cual es el producto con el cual oferta la compañía.
- Implementar un sistema de control eficiente que permita monitorear la información de cartera y trámite
- Elaborar un modelo de presupuesto de construcción basado en información histórica que proporcione costos fijos que no impacten las pérdidas y ganancias de la organización
- Estima el precio de venta de una propiedad estableciendo las expectativas del propietario sobre el precio de mercado por metro cuadrado

7.2.2. **Estrategias de innovación**

- Ofrecer a nuestros clientes finales la oportunidad de mantener su propiedad con todo lo necesario para que puedan disfrutarla desde el día de la entrega. Para ello, se establecen asociaciones con reconocidas empresas de muebles, electrodomésticos y decoración para garantizar precios atractivos para los compradores de viviendas
- Desarrollamos proyectos diseñados para atraer Millennials con propuestas modernas y funcionales que satisfagan sus intereses y necesidades (home office, áreas comunes de ocio, mayor interacción social).

7.2.3. **Estrategia de mercadeo**

Centrándose en el concepto de las cuatro P propuesto por McCarthy en los años 60, explica todas las estrategias que afectan a los productos, precios, distribución y comunicación de la

compañía. Esta propuesta proporciona una comprensión práctica de la interrelación de las actividades de marketing, abordando las áreas de enfoque más importantes y sigue siendo relevante gracias a su practicidad y recordabilidad.

Asimismo, se resaltan las siguientes estrategias:

Producto

- Innovar en el diseño de los proyectos, dando apertura a proyectos más vanguardistas y menos conservadores
- Crear nueva demanda y desarrollar nuevos conceptos de productos, especialmente entre los jóvenes que se niegan a invertir debido a la volatilidad del sector inmobiliario

Precio

- Sondear de manera mensual el valor de m² que se oferta en el sector donde la empresa tiene presencia con el fin de ofrecer un valor competitivo generando distinción en el precio
- De la investigación de mercado extraemos el precio que los consumidores esperan para abrir los proyectos que el mercado busca
- La flexibilidad en los métodos de pago permite un mayor valor del capital y un menor valor del capital

Comunicación

- Generar pautas atractivas en las principales revistas inmobiliarias
- Impulsar campañas publicitarias que realicen la curva de experiencia con la que cuenta la empresa
- Contar con una mayor presencia y publicidad en redes sociales
- Potencializar los canales de distribución y usarlos de puente para tener promoción en los lugares donde se tiene presencia

Distribución

- Fortalecer alianzas con las empresas en el sector y desarrollar un canal de ventas en línea directas para colombianos en el exterior
- Aprovechar el programa de Mi Casa Ya, subsidio a la tasa de interés para potencializar las ventas de las viviendas superiores a VIS el cual se oferta en la ciudad de estudio
- Desarrollar alianzas-público privadas enfocadas en combatir el déficit de viviendas en la ciudad.

Considerando lo anterior y el impacto negativo que tiene en la dinámica de ventas, se debe centrar todos tus esfuerzos en los clientes. Esto significa un cambio en la estructura de la cultura organizacional, la cual, requiere una cultura enfocada en el servicio al cliente. Esto requiere cambiar muchos procesos actuales para crear procesos más simples y mejorar la experiencia del cliente final.

La compañía tiene presencia en todo el país, lo que representa una ventaja competitiva, sin embargo, esta calidad tiene un gran impacto en el centro de mercadeo y ventas. Por ende, la constructora no toma en cuenta los aspectos culturales y genera actividades comerciales que no se ajustan a las realidades del mercado inmobiliario.

7.3.Objetivos a corto plazo

- Garantizar la participación del área de mercadeo y ventas de la compañía mediante en análisis de datos desde la etapa inicial de la consecución del proyecto
- Incrementar la participación en ventas en el mercado inmobiliario y su área de influencia
- Implementar nuevos canales de distribución que permitan maximizar las ventas y un adecuado posicionamiento en el mercado

- Posicionar el nombre de la compañía dentro de las primeras tres constructoras más reconocidas del país

Estrategia 1

La compañía debe replantear el “procedimiento de proyectos de construcción” el cual dicta que las áreas involucradas en el proceso de diseño desde su concepción hasta su legalización y aprobación en Curaduría es el presidente, Gerente General, Gerente de Diseño, Gerente de Proyecto, Directos CPC y Gerente de sucursal. En este procedimiento no se considera la participación del área de Mercadeo y ventas, por lo tanto, se plantea modificar el Procedimiento P-ARQ-001 pasando de esta manera a la versión número 25 la cual estará estructurada de la siguiente manera.

Tabla 30. Actividades de la compañía

Actividad	Responsable
La junta ejecutiva y/o el equipo directivo pasan la información de presentación del proyecto al departamento de diseño, marketing y ventas de la empresa, el cual define de forma colaborativa las características arquitectónicas del proyecto en función del precio de venta estimado y la capacidad de su organización para cumplir con requisitos de diseño como coste, tipo de proyecto, etc., y si se ha aplicado investigación de mercado o análisis competitivo.	<ul style="list-style-type: none"> • Gerente de diseño • Director de diseño sucursal • Presidencia • Fuerza Comercial
El plan de diseño se realiza en formato digital para planos de proyectos comunes, incluyendo la coordinación de diseño externo y reuniones de	<ul style="list-style-type: none"> • Arquitecto diseñador

ingeniería concurrentes en el plan, enviar el cronograma a Planificación Nacional y generar el cronograma del proyecto en el sistema JD.	<ul style="list-style-type: none"> • Director de diseño • Gerente de diseño
Desarrollan los proyectos necesarios teniendo en cuenta: - Requisitos de inclusión - Concepto de normas urbanísticas - Estudio de tipos y líneas de circunvalación - Tipología. - Número de casas. - Número de plazas de aparcamiento - Información comercial	Arquitecto diseñador
Diseños generales de arquitectura y urbanismo (últimas versiones de planos de planta, sección y fachada), con envío original al Administrador y copia al jefe de Gestión y Distribución de Planos Sucursales, jefe de Obra y CPC. La fábrica contratada es responsable.	<ul style="list-style-type: none"> • Arquitecto diseñador • Gerente sucursal • Director comercial
Inicia procesos de comercialización y ventas del proyecto	Fuerza comercial

Fuente: Elaboración propia

Se contrata una empresa que se especializa en investigar el mercado inmobiliario y su impacto, el cual brindará los datos más relevantes en un informe mensual que se podrá discutir y analizar con todo tu equipo de ventas al final del mes. Su objetivo es reforzar el conocimiento de las realidades del mercado por parte del sector de marketing y ventas, lo que, sumado a la experiencia del equipo comercial, les permitirá realizar una contribución significativa en el proceso de puesta en marcha de nuevos proyectos.

Estrategia 2

- Diversificar el portafolio de producto el cual complemente la oferta en la ciudad
- Implementar políticas más flexibles para el plan de pagos donde la fuerza comercial tenga más poder de negociación
- Desarrollar un enfoque de servicio al cliente interno y externo que permita la satisfacción del cliente final

Tabla 31. Plan de acción estrategia 2

ESTRATEGIAS	ACTIVIDADES	CRONOGRAMA						INVERSIONES	RECURSOS	RESPONSABLE
		BIMESTRE								
		ENE-DIC 2023								
1	2	3	4	5	6					
Diversificar Portafolio	Estudio de mercado enfocado a la demanda no atendida							\$ 6.000.000	Contratación firma especializada	Gerente de mercadeo y ventas
	Promover alianzas público-privadas para desarrollo de proyectos de gran envergadura							\$ 7.000.000	Tiempo de personal	Gerente de Arquitectura
Mayor negociación del área comercial	Establecer formas de pago más flexibles, estandarizadas que no generen pérdidas para la empresa							\$ 2.000.000	Tiempo de personal	Gerente financiero

Desarrollar un enfoque de servicio al cliente tanto interno como externo dentro de la compañía	Formar toda la estructura organizacional en servicio al cliente						\$ 10.000.000	Tiempo de personal	Gerencia y Recursos Humanos
	Capacitaciones						\$ 8.000.000	Tiempo de personal	Gerencia y Recursos Humanos
	Plan de incentivos						\$ 1.800.000	Tiempo de personal	Gerencia y Recursos Humanos
TOTAL							\$ 34.800.000		

Fuente: Elaboración propia

Estrategia 3.

- Establecer alianzas con distintas empresas del sector
- Desarrollar un canal de ventas directa para colombianos residentes en el exterior

Tabla 32. Plan de Acción estrategia 3

ESTRATEGIAS	ACTIVIDADES	CRONOGRAMA						INVERSIONES	RECURSOS	RESPONSABLE
		BIMESTRE ENE-DIC								
		2023								
1	2	3	4	5	6					
Fortalecer alianzas con distintas empresas en el sector	Suministrar capacitaciones al personal de las alianzas y organizaciones adscritas							\$ 5.000.000	Tiempo de personal	Gerente de mercadeo y ventas/Gerente de talento humano
	Establecer una tabla de comisiones que incremente el nivel de ventas mensuales del convenio							\$ 2.000.000	Tiempo de personal	Gerente financiero/Gerente de mercadeo y ventas
Desarrollar canal de venta directa	Establecer la estructura requerida para dar soporte a las ventas en el exterior							\$ 2.000.000	Tiempo de personal	Gerente financiero/Gerente de mercadeo y ventas
	Vincular a las personas encargadas de las ventas en el exterior							\$ 20.000.000	Tiempo de personal	Gerente de mercadeo y ventas/Gerente de talento humano
TOTAL								\$ 29.000.000		

Fuente: Elaboración propia

Estrategia 4

Trabajar el posicionamiento y reconocimiento de la compañía como una constructora que construye con calidad y que entiende las necesidades y expectativas de los clientes.

Tabla 33. plan de acción estrategia 4

ESTRATEGIAS	ACTIVIDADES	CRONOGRAMA						INVERSIONES	RECURSOS	RESPONSABLE
		BIMESTRE ENE-DIC 2023								
		1	2	3	4	5	6			
Posicionar la marca de la compañía	Definir el mercado objetivo							\$ 2.000.000	Tiempo de personal	Gerente de mercadeo y ventas
	Maximizar la utilización de redes sociales y medios publicitarios masivos							\$ 40.000.000	Tiempo de personal + pautas impresas y digitales	Gerente de mercadeo y ventas
Lograr el reconocimiento en el mercado por sus construcciones con altos	Generar planes de divulgación de los reconocimientos y certificaciones de calidad con la que cuenta la empresa							\$ 4.000.000	Tiempo de personal	Gerente de mercadeo y ventas

8. Plan y recomendaciones de implementación y aplicación

El modelo inicialmente se encuentra sesgado ya que la base de datos compartida por la compañía está compuesta por compradores de inmuebles, el cual no tenemos información sobre las personas que no compraron algún inmueble, personas que visitaron las salas de ventas o realizaron solo la cotización de algún inmueble y no se culminó la compra, esto nos ayudaría a poder analizar las características y gustos en común, para poder abordar esa población que desiste en el proceso de adquisición.

Es de gran importancia tener los datos estructurados teniendo en cuenta todos los campos diligenciados sin dejar campos vacíos, puesto que esto evita que en nuestra base se encuentren datos nulos, eliminando información de la base de datos al realizar la limpieza, por lo cual se puede determinar que no se tienen los procesos claros para administrar los datos de forma estructurada.

Se debe generar procedimientos claros en el área comercial para que la compañía obtenga la información necesaria y completa, para realizar los análisis del comportamiento de sus clientes, este procedimiento permite validar que la información capturada en la base de datos sea con calidad, relevancia, unicidad, completitud, correctitud, confiabilidad, integridad, precisión y consistencia.

Se debe crear procedimientos que incluya políticas y responsabilidades desde la creación de los datos, hasta el almacenamiento y administración de estos creando un gobierno de datos el cual se puede encargar de velar por la calidad de los datos para que nuestros análisis no se vean sesgados por la ausencia de la información.

Se recomienda capacitar al personal para realizar las funciones que vaya acorde a los procedimientos que se plantean para que los datos se encaminen en la mejora continua, de esta manera tendremos unos datos mas confiables, darles a conocer al personal la importancia de

complementar toda la información cuando se realiza una venta, ya que el recurso mas importante de toda organización es la información.

9. Conclusiones

La compañía es una empresa bien organizada, como lo demuestran los certificados internacionales que posee, por ende, la elaboración y utilización de procesos internos es una de las fortalezas de la organización, debido a que el comportamiento de los empleados está ligado a la empresa, por lo que es necesario utilizar los acuerdos a los que se ha llegado para que sea cada vez más sólida, creando oportunidades para capacitar, apoyar y aumentar el conocimiento de los empleados para que puedan disfrutar de este poder, se llama cultura de la organización.

Asimismo, la investigación realizada sobre la herramienta diseñada para la identificación de género enmarcó un gran trabajo logrando poder identificar el género al que pertenece el comprador basándose el código con el nombre. El objetivo de la compañía de realizar una segmentación de clientes genera que se creen campañas dirigidas al público adecuado minimizando el posible rechazo de las campañas que se envíen por parte de la compañía.

El modelo de segmentación se generaron varios escenarios, los cuales se agruparon según el posible evento que se podría dar, analizándose desde la perspectiva de persona, inmueble y compra. Esta información puede ser utilizada para la toma de decisiones, el mejoramiento de los procesos y aumento de eficiencia en los recursos.

Con un buen gobierno de datos se puede generar modelos de seguimiento, control de indicadores y trazabilidad del dato, podemos crear data warehouse consistente para la creación de tableros de control y generar plataformas de visualización inteligentes de cifras.

Por último, para garantizar que los clústeres sean significativos y útiles para el análisis, es importante evaluar la calidad de su segmentación utilizando métricas internas y externas. En resumen, la segmentación K-Means es un enfoque poderoso para examinar la estructura de los datos y encontrar patrones significativos. Sin embargo, para obtener resultados confiables y

significativos, es importante ser consciente de sus limitaciones y considerar cuidadosamente la selección de parámetros y el preprocesamiento de datos.

Referencias bibliográficas

- Alayon, R. A., Moncada, D. A., Medina, V. H., & Rodríguez, J. E. (s/f). *Estimación de rendimiento académico a través de técnicas para minería de datos.*
- Alberto, C., Cardona, E., & Saint-Priest Velásquez, Y. (s/f). *Characterization model of customer profile adapted to building sector with data mining: case study building company.*
- Alejandra, P., Camargo, G., De, U., Andes, L., & De Ingeniería, F. (s/f). *IMPLEMENTACIÓN DE UN MODELO MACHINE LEARNING PARA LA ESTIMACIÓN DEL VALOR DEL METRO CUADRADO DE UN INMUEBLE UBICADO EN CUNDINAMARCA.*
- Business Intelligence aplicado al área de compras en la empresa. (s/f). *Frank Oscar Gamarra Valencia.*
- Cesar Moyano-Alulema, J. I. (s/f). *Modelo de gestión por procesos para la comercialización de bienes inmuebles en urbanizaciones Ciencias económicas y empresariales Artículo de investigación.* 6, 395–417. <https://doi.org/10.23857/dc.v6i3.1225>
- Ghazzawi, A., & Alharbi, B. (2019). Analysis of Customer Complaints Data using Data Mining Techniques. *Procedia Computer Science*, 163, 62–69.
<https://doi.org/10.1016/j.procs.2019.12.087>
- Haya Pablo. (2021). <https://www.iic.uam.es/innovacion/metodologia-crisp-dm-ciencia-de-datos/>.
- Kaur, p., & Singh, M. (2023). *Modelo de segmentación de clientes usando K-means.*
- Kyocera. (2017). *Diferencia entre datos estructurados y no estructurados.*
<https://www.kyoceradocumentsolutions.es/es/smarter-workspaces/insights-hub/articles/datos-de-personas-como-conseguirlos.html#:~:text=En%20resumen%2C%20mientras%20que%20los,de%20objetos%20valiosos%20pero%20desordenados.>

La necesidad de la evolución de la gerencia en el sector de la construcción en Colombia en la cuarta revolución industrial. (s/f).

Marzal Varó, A., & Gracia Luengo, I. (2009). *Introducción a la programación con Python.*

Universitat Jaume I. Servei de Comunicació i Publicacions.

Ncr, (, & Clinton, J. (2000). *Step-by-step data mining guide.* DaimlerChrysler.

San, A., & Cabrero, J. (s/f). *El futuro de la valoración inmobiliaria Big Data y modelos AVM.*

Sociedad, U. Y., Jorge, M., Rivero Pérez, L., Yaimara, D. C., Santana, P., Harenton, I. P., &

López, M. (2016). *Volumen 8 / Número 2 / Mayo-agosto.* <http://rus.ucf.edu.cu/>

Vázquez Rodríguez, S., Sergio, L., & Torres, E. (s/f). *LOS MERCADOS INMOBILIARIOS EN LA ÉPOCA DEL BIG DATA.*

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, *1*(1), 3–14.

<https://doi.org/10.1111/j.2041-210x.2009.00001.x>