

Formulación de modelos clásicos y de aprendizaje de máquina para el pronóstico del precio de  
los proyectos de construcción

Gina Katherine Tellez Guevara

Director

Germán Arcesio Ortiz Basto

Universidad de la Sabana

Maestría en Gerencia en Ingeniería

Bogotá

2022

## **Dedicatoria**

A Dios por acompañarme y darme la oportunidad de crecer en todos los aspectos de mi vida. A mis padres y familia por todas sus lecciones, enseñanzas, cariño y apoyo en este largo camino. A mi tutor por ser un excelente guía en este proceso educativo.

## **Agradecimientos**

Agradezco a Dios por permitirme recorrer los caminos del conocimiento contribuyendo con mi crecimiento profesional.

A mis padres por su apoyo incondicional, enseñanzas y por inculcarme el deseo de superación que llevo tan arraigado que me ha permitido ir cumpliendo una a una todas las metas, objetivos y sueños que me he propuesto.

A la Universidad y a mi tutor por brindarme todo su conocimiento y experiencia.

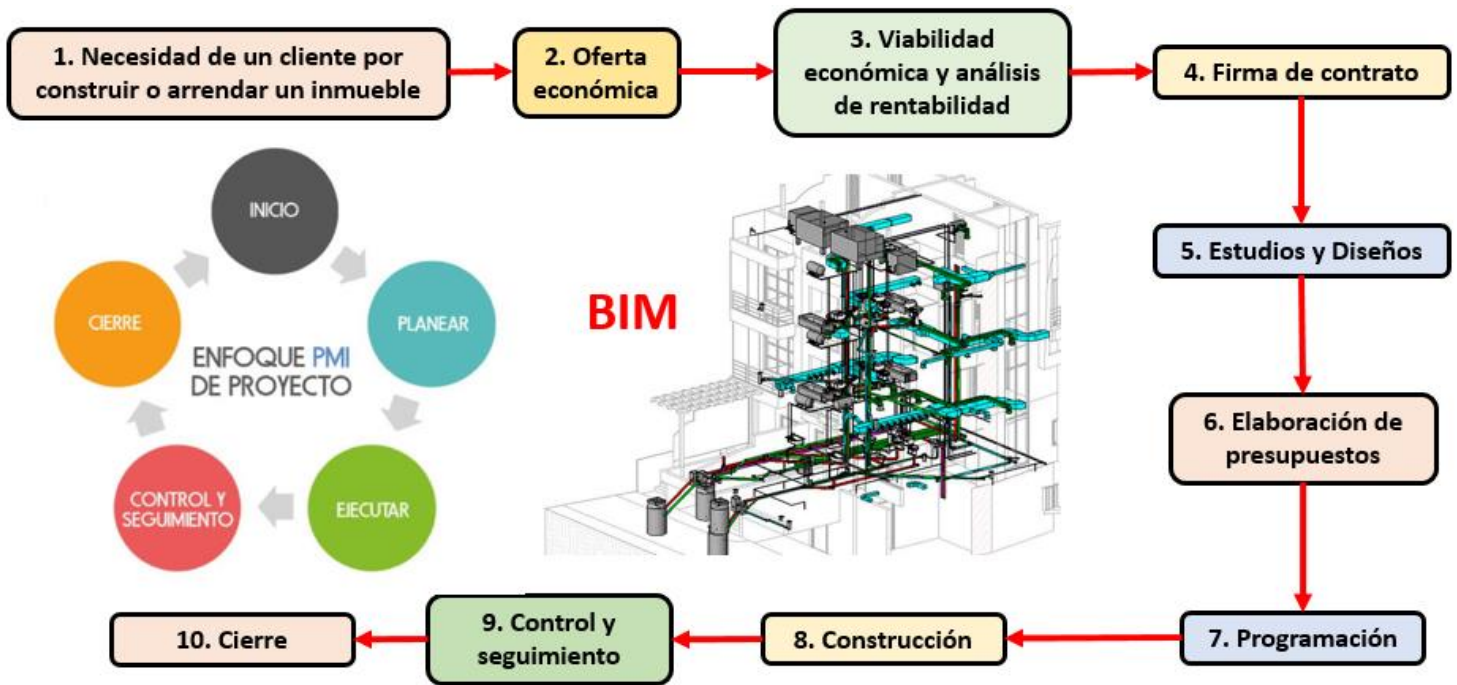
## Tabla de contenido

<b>1. Introducción .....</b>	<b>5</b>
<b>2. Marco teórico .....</b>	<b>8</b>
<b>4. Metodología .....</b>	<b>14</b>
<b>5. Resultados.....</b>	<b>15</b>
Proceso de recolección de información y organización de la base de datos .....	15
Identificación de las variables de entrada y salida que se van a utilizar en cada modelo .....	25
Modelos supervisados .....	27
Ajuste del modelo de regresión lineal múltiple .....	27
Ajuste del modelo de árboles de decisión.....	29
Ajuste por máquina de soporte vectorial .....	30
Pronóstico mediante redes neuronales .....	32
<b>6. Conclusiones y recomendaciones.....</b>	<b>36</b>
<b>7. Referencias bibliográficas .....</b>	<b>38</b>
<b>9. Anexos .....</b>	<b>44</b>
9.1 Anexo 1 - (Variables categóricas Ciudad y Zona) - Base de datos General .....	44
9.2 Anexo 2: Niveles (Variables categóricas Estrato y Barrio) - Base de datos General.....	47

# 1. Introducción

La ejecución de proyectos por parte de empresas del sector de la construcción tiene un proceso que en la mayoría de los casos inicia con la necesidad presente en el cliente por construir o arrendar un inmueble, este interés conlleva a que se elabore una oferta económica (ver figura 1) por parte de la compañía constructora; posterior a esto, se procede a realizar la viabilidad económica y el análisis de rentabilidad, luego se firma el contrato, se realizan estudios y diseños, se elabora el presupuesto, se planea la programación, se construye, el siguiente paso que es control y seguimiento, para finalmente llegar al cierre, tal como se presenta en la figura 1.

Figura 1. Proceso de negociación proyectos de construcción



Fuente: elaboración propia del autor

Algunas compañías del sector están presentando como problemática que el precio que se entrega a los clientes por medio de una oferta económica durante la etapa de planeación se realiza sin estudios ni diseños definitivos basándose únicamente en la experiencia del profesional encargado, lo anterior puede ocasionar altas desviaciones en el precio obtenido a partir de los costos y tiempos largos en la generación de las ofertas, situación que obstaculiza la toma de decisiones informadas para que el proyecto sea viable y rentable. Lo anterior es importante porque es frecuente que existan cambios significativos entre los precios estimados en la etapa de planeación en comparación a lo

invertido en la etapa de entrega del proyecto.

“En el análisis de tiempos y precios, el factor de planeación resulta comúnmente significativo en la variación, lo que se traduce en que la planeación deficiente repercute en los proyectos generando diferencias de tiempo y precio” (Lozano et al, 2018, p. 136). La incidencia de la planeación en los precios del proyecto se evidencia mejor en la tabla 1.

Tabla 1 Factores influyentes en las variaciones de los precios

No.	Grupo al que pertenece	Factor	% varianza
1	Planeación	Falta de materiales	28 %
2	Partes involucradas	Falta de comunicación entre las partes	9 %
3	Agentes externos	Fluctuaciones de la moneda	7 %
4	Agentes externos	Prácticas fraudulentas	7 %
5	Otros aspectos	Tipo de proyecto	6 %
6	Partes involucradas	Productividad laboral inadecuada	5 %
7	Planeación	Cambios en el alcance del contrato	4 %
8	Otros aspectos	Experiencia del entrevistado	3 %
9	Planeación	Planeación de cronograma inadecuada	3 %
10	Planeación	Falta de maquinaria	3 %

Fuente: Lozano et. all. (2018).

Los datos mostrados en la tabla 1 son de vital importancia en la etapa inicial dado que “con una buena planeación de la obra se logra determinar, en primera instancia, el equipo más adecuado en cuanto a operación y precio” (Rivera, 2015, p.2). Además de la planeación, también se presentan agentes externos que inciden en un precio inviable del proyecto.

No obstante, la realidad en el contexto de la mayoría de las empresas del sector es que la decisión del precio se realiza de forma rápida con el objetivo de aumentar las oportunidades de lograr una negociación y a un precio competitivo para el cliente que incida en la decisión de aprobación. Sin embargo, para Camargo (2015) la elección del precio de los proyectos de construcción debe ser consecuencia de un ejercicio cauteloso debido a que existen numerosas variables que afectan el precio como la ubicación, características del diseño, tiempo de ejecución, materias primas entre otros. “El mayor grado de incertidumbre por riesgo se encuentra en las primeras etapas del proyecto y que las decisiones en dicha etapa tienen un alto impacto en el costo y tiempo finales” (Camargo, 2015, p. 17).

A la anterior situación se suma la dificultad de contar con las suficientes herramientas tecnológicas para establecer el precio del proyecto, ya que en la mayoría de los casos es un proceso realizado por un profesional apoyado en el programa Excel, situación que aumenta los riesgos de establecer

un precio incorrecto debido a causas como “falta de conocimiento, exceso de confianza, mala planeación, entre otras. (...) Se pueden presentar errores que traerían consecuencias en costos financieros como también en vidas humanas” (García, Hincapié y Pérez, 2015, p. 13).

Al respecto, se debe tener presente que los métodos tradicionales se ven obstaculizados para estimar con precisión costos del proyecto debido a la gran cantidad de variables significativas y sus interacciones. Por lo tanto, los métodos tradicionales tienen una aplicabilidad limitada.

Muchos contratistas preparan cotizaciones basadas en intuiciones, sin estimar los costos del proyecto, de modo tal que las ofertas resultan demasiado altas y pierden el contrato, o mucho peor, obtienen el trabajo, a un precio que está muy por debajo del costo (OIT, 2018, p. 2).

Lo anterior según Potes y Vargas (2017) conlleva a que los proyectos presenten retrasos y sobrecostos que afectan la satisfacción del cliente, la situación financiera de la empresa y la capacidad operativa en el corto, mediano y largo plazo.

Entre otras problemáticas, se encuentra la dependencia de la empresa con relación al profesional a cargo de la responsabilidad de estimar el precio de los proyectos; también existe una disponibilidad limitada de información durante las primeras etapas; y en la estimación prevalece la incertidumbre y vaguedad debido a los datos incompletos y complejos.

A partir de los anteriores argumentos se identificó la necesidad de comprender la dinámica de los precios en los proyectos de construcción a partir de modelos supervisados como apoyo en la toma de decisiones gerenciales en etapas de pre factibilidad y análisis de viabilidad económica que permita realizar predicciones con respecto a los precios de proyectos de construcción en la empresa Zona Franca Bogotá; durante el desarrollo del estudio se identificó que los datos obtenidos estaban disponibles con información sobre los precios, sin embargo, se tuvo el obstáculo que la base de datos de la empresa en mención no era suficiente para poder alimentar el modelo, por lo que se tuvo que realizar con una población más amplia que fueron los precios de proyectos de construcción en Cundinamarca suministrados por la Cámara Colombiana de la Construcción (Camacol) desde el año 1995 al 2020.

La investigación posibilitará que las compañías del sector puedan cuantificar los precios con mayor confianza e identificar un presupuesto para brindar una oferta competitiva, y poder así reducir los riesgos de errores en precios que conlleven a retrasos, sobrecostos y afectaciones en la satisfacción de los clientes y reputación de las organizaciones.

En consecuencia, las empresas del sector de la construcción podrán ser eficientes y eficaces, lo cual incide en la competitividad y productividad. Si las organizaciones logran una mayor capacidad en la producción con costos menores, entonces “se pueden establecer precios de mercado más bajos y esto redundará en una unidad productiva más competitiva respecto al resto de agentes del mercado” (Camacol, 2018, p. 10), además de alta satisfacción en los clientes, buena reputación y posicionamiento en el sector.

**Pregunta de investigación:** ¿Usando modelos supervisados se podría predecir el precio de un proyecto de construcción en etapa de pre factibilidad?

## 2. Marco teórico

- Etapa de planeación proyectos de construcción

La ejecución de proyectos y obras de construcción implica el cumplimiento de unas etapas las cuales son el inicio en donde se establece la misión, visión, objetivos y justificación del proyecto; luego está la planeación que corresponde a establecer las estrategias que aportarán al cumplimiento de los objetivos, brindando orientación sobre el cómo y cuáles pasos se deben realizar; posterior a esto se encuentra la ejecución donde se implementa el plan, por lo que se ejecutan las acciones; luego se debe realizar seguimiento y control, proceso en donde se compara lo ejecutado con relación a lo que se planeó (Mora, 2015).

Según la Guía de los fundamentos para la Dirección de Proyectos (2008) como parte de la planeación, algunas de las principales actividades son el desarrollo del plan para la dirección del proyecto, definir el alcance, estimar los recursos y costos, elaborar el cronograma, determinar el presupuesto, identificar los riesgos y planear su gestión, entre otros.

Al respecto, se conoce que la etapa de planeación es una de las más importantes que incide en la toma de decisiones sobre pre factibilidad y viabilidad económica, esto debido a que la estimación de los costos y el presupuesto final es un ejercicio que se realiza en esta etapa (Mora, 2015).



De forma complementaria, Gaviria (2016) afirma que:

Para realizar dicha planeación el papel que desempeña el análisis de costos se convierte en una herramienta necesaria para lograr obtener la información pertinente, donde se permite visualizar el costo final proyectado de cada obra en ejecución y planear la facturación anualmente (p. 13).

Así mismo, el conocimiento de los costos permitirá establecer el precio final del proyecto. Gaviria (2016) indica que el precio unitario depende del costo directo (mano de obra, materiales, maquinaria y equipo); el costo indirecto, es decir todos los gastos importantes para conservar la ejecución del proyecto, por ejemplo la supervisión, los salarios de los trabajadores administrativos, entre otros. También se conoce que como parte de los costos indirectos se encuentran:

Indirectos de campo que se calculan mediante la obtención del porcentaje que representa el presupuesto de gastos de campo en relación con el costo directo total del proyecto y de oficina central, el cual resulta de prorratear el presupuesto proforma de la oficina central entre la suma total de los costos directos de obra que la empresa espera realizar en el periodo analizado por el presupuesto y otros, los cuales se desglosan en imprevistos, gastos de financiamiento y utilidad (López de Ortigosa, 2010, p. 55).

Otro de los aportes es de Peña (2009) para quien la etapa de planeación es indispensable para tomar decisiones de rentabilidad económica, esto debido a que es necesario establecer el nivel de inversión, el periodo de vida útil que tiene el proyecto, los costos tanto de producción como materias primas, los insumos, manos de obra, fuentes de financiación, precio, entre otros.

La importancia que sustenta Peña (2009) sobre los precios integra indicadores y variables macroeconómicas, esto debido a cambios en el:

Valor del dinero en el tiempo, la tasa de inflación nacional, la tasa de devaluación monetaria, el nivel de incertidumbre y riesgo de la inversión, las políticas económicas del gobierno, las políticas de precios insumos/productos, políticas monetarias y crédito y las políticas arancelaria y de comercio exterior (p. 17).

- Aprendizaje de máquina

El aprendizaje de máquina es comprendido como un área que se enfoca en estudiar cómo construir programas de computadores que contribuyan a mejorar el desempeño de alguna tarea, apoyado en la experiencia. Para lograrlo, tienen en cuenta la teoría de la inteligencia artificial, estadística, probabilidad, psicología, neurobiología, teoría de control y complejidad computacional, entre otros (Mitchell, 1997).

De forma complementaria, se debe tener presente que para que se logre el aprendizaje es necesario que existan decisiones sobre el tipo de entrenamiento, la función objetiva a ser aprendida, la representación y el algoritmo que permitirá aprender esa función a través de ejemplos de entrenamiento (Mitchell, 1997).

Al respecto, se debe tener presente que los algoritmos de aprendizaje han sido considerados como efectivos y útiles para trabajar con grandes bases de datos que tienen regularidades, por lo que pueden ser descubiertas desde la automatización y “en dominios que son poco entendidos y donde los humanos no poseen el conocimiento necesario para desarrollar algoritmos efectivos, y en dominios donde los programas se deben adaptar en forma dinámica para responder a condiciones cambiantes en el ambiente” (Godoy, 2014, p. 105).

Este proceso de aprendizaje de máquina es importante para Gutiérrez y Rodríguez (2018) porque el rápido crecimiento de la información ha aumentado la cantidad de datos que el ser humano debe tener en cuenta para la toma de decisiones, hasta el punto de ser imposible de procesar y analizar, es en esas situaciones en donde las técnicas como el aprendizaje de máquina integran modelos que posibilitan el análisis de grandes volúmenes de datos, para que sea un proceso rápido, preciso y confiable. “El resultado de usar aprendizaje de máquina son predicciones de alto valor que pueden guiar y tomar las mejores decisiones sin necesidad de la intervención humana” (Gutiérrez y Rodríguez, 2018, p.3).

Tal como lo menciona Raschaka y Mirjalili (2017) hay tres tipos de aprendizaje automático: Aprendizaje supervisado, aprendizaje no supervisado y el aprendizaje reforzado. “El objetivo del aprendizaje supervisado es aprender de un modelo, a partir de datos de entrenamiento etiquetados, que nos permite hacer predicciones sobre datos futuros o no vistos. Aquí el termino supervisado se refiere a un conjunto de muestras donde las señales de salida deseadas (etiquetas) ya se

conocen”. Los árboles de decisión, calificación de Naive Bayes, regresión por mínimos cuadrados y máquinas de soporte vectorial hacen parte de los algoritmos más habituales en los métodos supervisados tal como lo menciona Paloma Recuerdo de los Santos (2021). Por su parte las redes neuronales pueden ser algoritmos supervisados o no supervisados tal como lo menciona Rodrigo Salas (2004).

- Redes Neuronales Artificiales

La red neuronal según Galán y Martínez (2018) es una representación abstraída del modelo neuronal del cerebro humano, estas redes se forman a través de numerosos elementos simples que están interconectados. Algunas de las principales ventajas son el aprendizaje adaptativo, la auto-organización, tolerancia a fallos, acción en tiempo real y facilidad de uso.

En definitiva, las redes neuronales no son más que un modelo artificial y simplificado del cerebro humano, que es el ejemplo más perfecto del que disponemos para un sistema que es capaz de adquirir conocimiento a través de la experiencia (Matich, 2001, p.4).

Las redes neuronales artificiales permiten clasificar o explicar un valor a partir de un conjunto de datos observados. En Rodrigo Salas (2004) se especifica el uso de la suma ponderada de su entrada y luego la aplicación de una función de activación para obtener una señal que será transmitida a la próxima neurona en su funcionamiento, y a través de un algoritmo de aprendizaje supervisado o no supervisado, las redes neuronales ajustan su arquitectura y parámetros minimizando alguna función de error.

- Modelo de regresión lineal

El modelo de regresión lineal consiste en explicar la relación existente entre una variable que es dependiente (variable respuesta), y otras que son independientes (variables explicativas). En otras palabras, se busca describir cómo influyen la variable “x” sobre la variable respuesta y.

X: variable independiente, explicativa o exógena

Y: variable dependiente, respuesta o endógena

A continuación se presenta la estructura del modelo de regresión lineal:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

$$X_1, X_2, \dots, X_k$$

A partir de la anterior expresión es posible conocer el comportamiento de la variable respuesta “y” a partir de dos conjuntos de variables, uno que integra las variables explicativas  $X_1, X_2, \dots, X_k$  y otro que corresponde al error “e” que no se puede controlar como la perturbación o error aleatorio y “que provoca que la dependencia entre las variables dependiente e independiente no sea perfecta, sino que esté sujeta a incertidumbre” (Universidad de Jaén, 2015, p.3)

- Inteligencia Artificial en el sector de la construcción

Según Dallasega, Rauch, & Linder (2018) la cuarta revolución industrial (Industria 4.0) está cambiando no solo la industria manufacturera, sino también la industria de la construcción conocida como Construcción 4.0. Este término tiene muchas definiciones generales por ejemplo T. Bock, 2015) muy generalmente propuso una definición de "automatización de la construcción" como un nuevo conjunto de tecnologías y procesos que cambiarán todo el curso y la idea de construcción de una manera fundamental. (K. Jung, B. Chu, 2013) usó una definición más exacta, refiriéndose a la "automatización de la construcción" como una tecnología de fábrica de construcción centrada en la máquina para aplicar sistemas robóticos en el campo de la construcción.

Los países que aportan el mayor número de publicaciones científicas e investigaciones con respecto a la automatización en el sector de la construcción se presentan en la siguiente tabla:

Tabla 2 Porcentaje de los 10 países principales (Oesterreich & Teuteberg, 2016)

Rank Country	Scientific publications		Practical publications		Total	
	Adjusted #	Adjusted %	Adjusted #	Adjusted %	Adjusted #	Adjusted %
1 USA	32.1	21.5%	64.0	53.8%	96.1	35.9%
2 UK	8.6	5.8%	18.0	15.1%	26.6	9.9%
3 China / Hongkong	19.7	13.2%	1.0	0.8%	20.7	7.7%
4 Canada	15.7	10.5%	3.0	2.5%	18.7	7.0%
5 Australia	7.8	5.2%	7.0	5.9%	14.8	5.5%
6 Korea	12.3	8.2%	2.0	1.7%	14.3	5.3%
7 Germany	4.8	3.2%	3.0	2.5%	7.8	2.9%
8 Malaysia	3.3	2.2%	3.0	2.5%	6.3	2.3%
9 Sweden	3.0	2.0%	2.0	1.7%	5.0	1.9%
10 Taiwan	4.8	3.2%		0.0%	4.8	1.8%

Estas publicaciones han sido clasificadas por grupos o clúster de investigación tal como se presenta en la siguiente tabla:

Tabla 3 Clasificación por grupos o clúster (Oesterreich & Teuteberg, 2016)

Cluster	Key technologies and concepts in the context of Industrv 4.0	Number of relevant scientific publications	Number of relevant practical publications
Smart Factory (C1)	Cyber-Physical systems (CPS)/ Embedded systems	2	0
	Radio-Frequency Identification (RFID)	11	6
	Internet of Things (IoT)/ Internet of Services (IoS)	0	27
	Automation	21	3
	Modularisation/ Prefabrication	7	4
	Additive Manufacturing	0	8
	Product-Lifecycle-Management (PLM)	2	3
	Robotics	6	13
Human-Computer Interaction (HCI)	2	1	
Simulation and modelling (C2)	Simulation tools/ Simulation models	55	1
	Building Information Modelling (BIM)	30	11
	Augmented/Virtual/Mixed Reality (AR/VR/MR)	18	13
Digitisation and virtualisation (C3)	Cloud Computing	6	5
	Big Data	0	9
	Mobile Computing	2	1
	Social Media	3	6
	Digitisation	4	8
	<b>Total</b>	<b>169</b>	<b>119</b>

De lo anterior se establece que en el primer grupo Smart Factory contiene una amplia gama de tecnologías y conceptos para automatizar el proceso de construcción y crear una “Fábrica Inteligente” para el entorno de la construcción. En general, hay algunos enfoques interesantes para crear una "construcción inteligente en sitio". Por ejemplo, (Akanmu & Anumba, 2015) se ocupa de la integración de Cyber-Physical Systems (CPS) con el fin de facilitar la gestión bidireccional, coordinación entre modelos virtuales y construcción física, incluida la validación de la prueba de concepto desarrollada con sistemas prototipo por profesionales de la industria. En lo que se refiere a tecnología de sensores, Radio-Frequency identificación (RFID) ofrece múltiples soluciones para la automatización del proceso de construcción, permite al personal del sitio efectivamente rastrear y administrar sus activos como herramientas, equipos, materiales y componentes de construcción para la optimización, cronograma del proyecto y costo del proyecto (Majrouhi Sardroud, 2012). Otro concepto central en este grupo es el Internet de las cosas y Servicios (IoT y IoS), que permite la creación de aplicaciones virtuales o redes para soportar un entorno de fábrica inteligente.

### 3. Objetivos

#### Objetivo general

Pronosticar el precio total de los proyectos de construcción en etapa de pre factibilidad mediante el ajuste de modelos supervisados que permitan una mayor precisión en la oferta

económica.

### **Objetivos específicos**

- Describir el proceso de recolección de información y organización de la base de datos sobre los precios de los proyectos de construcción.
- Identificar las variables de entrada y salida que se van a utilizar para el ajuste de los modelos que van a permitir el pronóstico del precio.
- Formular y evaluar modelos de aprendizaje de máquina que permita pronosticar el precio de un proyecto de construcción.
- Definir las medidas de desempeño o indicadores que permitan analizar el comportamiento de los modelos para identificar el que ajusta adecuadamente los datos para el pronóstico.

### **4. Metodología**

La investigación es cuantitativa debido a que tiene como objetivos medir fenómenos, utilizar la estadística, probar la hipótesis y la teoría, y se basa en planteamientos acotados. En cuanto al proceso, es deductivo, secuencial, probatorio y analiza la realidad objetiva (Hernández, Fernández y Baptista, 2016).

Como parte de sus bondades o ventajas en comparación a otros tipos de investigación, se encuentra que posibilita la generalización de resultados, controlar los fenómenos, obtener información precisa, la réplica y predicción (Hernández, Fernández y Baptista, 2016).

Para el desarrollo de un estudio cuantitativo, Tamayo (2007) indica que es necesario tener una muestra seleccionada de forma aleatoria o discriminada, pero sí debe ser representativa. En este proceso prevalece la aplicación del cuestionario para poder producir datos que sean analizados estadísticamente y que posibiliten verificar, aprobar o rechazar las relaciones entre las variables establecidas operacionalmente.

La técnica de recolección de información se basó en la revisión de documentos organizacionales que corresponde a registros oficiales de una organización. En este caso específico se trabajó con una base de datos del sector de construcción suministrada por la Cámara Colombiana de la Construcción (CAMACOL) del censo que realizan desde el año 1995 al 2020.

CAMACOL es una asociación gremial de carácter nacional sin ánimo de lucro que reúne a nivel nacional empresas y personas naturales relacionadas con la cadena de

valor de la construcción. Se creó en Medellín el 14 de septiembre de 1957 como iniciativa de un grupo de industriales y empresarios colombianos reunidos en la primera convención nacional de constructores. El fundamento para crear Camacol fue la necesidad de constituir una identidad que velara por los intereses de la industria de la construcción y que estuviera conformada por constructores, representantes de la industria y del comercio (Cámara Nacional de la Construcción, Camacol, s.f)

La base de datos consolidada para la investigación se depuró para los ajustes en los modelos planteados, donde se usaron programas de libre acceso R y Python usando R Studio como entorno de desarrollo de R y Google Colaboratory para la programación y ejecución de Python.

RStudio se define como un entorno de desarrollo integrado para el lenguaje de programación R dedicado a la computación estadística y gráficos (RStudio, s.f) y Google Colaboratory permite ejecutar y programar en Python (Colaboratory, s.f).

Python es un lenguaje de programación creado en 1980 por Guido Van Rossum en el Centro para las Matemáticas y la informática de los Países Bajos. Este es un lenguaje de programación sencillo de manejar el cual tiene como ventaja que es libre de licencia por lo que para su uso no se requiere pagar costos de derecho de autor (Giraldo, 2020)

En cuanto a la base de datos suministrada por CAMACOL, se caracteriza por tener 51.579 registros y 34 variables para la base de datos general y 21,190 registros con 124 variables para la base de datos complementaria de acabados, después de la depuración se logró tener 40.475 registros con 117 variables.

Así mismo, se menciona que el estudio es exploratorio debido a que se investiga sobre una problemática poco estudiada a nivel general de elaboración de proyectos empresariales en diferentes sectores incluyendo el de construcción. También se caracteriza porque se indaga desde un enfoque innovador, contribuye a identificar conceptos promisorios y preparan el terreno para la realización de futuras investigaciones (Hernández, Fernández y Baptista, 2016).

## **5. Resultados**

### **Proceso de recolección de información y organización de la base de datos:**

**Proceso de recolección de información:** La información con la que se entrenaron los modelos

fue tomada del censo mensual de edificaciones que se realiza en todas las regionales de Camacol desde el año 1995, con una cobertura de más de 120 municipios en edificaciones residenciales y no residenciales de todos los destinos y usos. Para la presente tesis se tomó una muestra en formato Excel para la Regional Cundinamarca en la cual se presenta 52.085 registros con 34 variables para la base de datos general y 21.190 registros con 124 variables para la base de datos de acabados. Las variables se presentan a continuación:

Tabla 4 Base de datos general

VARIABLES				
1. Longitud	2. Latitud	3. ID Proyecto	4. ID Etapa	5. Código del proyecto
6. Regional	7. Ciudad	8. Zona	9. Destino Etapa	10. Estrato
11. Barrio	12. Nombre del proyecto	13. Dirección	14. Sistema constructivo	15. Nombre de la etapa
16. Número de unidades por etapa	17. Tipo por etapa	18. Usos	19. Fecha inicio ventas	20. Fecha inicio Construcción
21. Fecha terminación construcción	22. Fecha de entrega	23. ID_Tipo	24. Segmento	25. Unidades por tipo
26. Área por tipo	27. Área Total	28. Alcobas	29. baños	30. Precio de Venta
31. Precio por metro	32. Localidad	33. Estado	34. Fase	

Fuente: Elaboración propia del autor

Tabla 5 Base de datos Acabados

VARIABLES				
1. Activo/Inactivo	2. ID Proyecto	3. Código de proyecto	4. Nombre proyecto	5. Regional
6. Ciudad	7. Dirección Proyecto	8. Teléfono	9. Destino	10. uso general del proyecto
11. zonainternet	12. zonabbq	13. zonabalcn	14. zonasesocial	15. zonaascensores
16. zonagimnasio	17. zonaszonasverdes	18. zonaguardera	19. zonacicloruta	20. zonapiscinas
21. zonabar	22. zonacanchamltiple	23. zonacanchasquash	24. zonacanchadefutbol	25. zonalobbyrepcin
26. zonaoficinaadministracin	27. zonaespejodeagua	28. zonaparqueinfantil	29. zonacitofona	30. zonaaireacondicionado
31. zonacinema	32. zonasenderodetrote	33. zonaterrazaskyclub	34. zonafuente	35. zonaredcontraincendios
36. zonazonadepesca	37. zonakiosco	38. zonacuartodebasuras	39. zonacanchatenis	40. zonasindefinir
41. zonashutbasuras	42. zonanaqueabastecimientodeagua	43. zonasalcomunal	44. zonacaldera	45. zonagolfito
46. zonaportera	47. zonasalndejugos	48. zonacampodegolf	49. zonalavandera	50. zonabicycletero
51. zonazonashumedas	52. zonasalonesespecialaladejunta	53. zonaplantaelctricadeemergencia	54. zonaplazoleta	55. zonacentrocmputo
56. zonacafeteria	57. zonarestaurante	58. zonaauditorios	59. zonacapilla	60. zonabateradebaos
61. zonavestier	62. zonabiblioteca	63. zonascuartogases	64. zonamascotas	65. zonacuartoequipos
66. zonadenios	67. zonaarealounge	68. zonasalastudio	69. zonaescaleras	70. zonaminimarket
71. zonaconjcerrado	72. zonasalamultiple	73. zonateatrino	74. zonacoworking	75. zonayoga
76. zonaspa	77. zonasasuna	78. zonadeposito	79. zonacentrocultural	80. zonateatro
81. zonaviadeacceso	82. zonanatquereservagua	83. zonarooftgarden	84. zonana	85. zonamuroescalar
86. zonanoresponde	87. placaentrepiso	88. sistema_constructivo	89. subestructura_cimentacion	90. subestructura_sotanos
91. divisiones_interiores	92. fachada	93. cubiertas	94. ventaneria	95. id_etapa
96. abiertacerrada	97. cdigoetapa	98. nombreetapa	99. unidadesetapa	100. destinoetapa
101. usoetapa	102. otrosusosetapa	103. pisos_alcobas	104. muros_alcobas	105. cielo_raso_alcobas



106. pisos_cocina	107. muros_cocina	108. cielo_raso_cocina	109. tipo_cocina	110. meson_cocina
111. muebles_cocina	112. dotacion_cocina	113. pisos_baños	114. muros_baños	115. cielo_raso_baños
116. meson_baños	117. division_baño	118. lavamanos	119. mueble_lavamanos	120. lavadero
121. puerta_principal	122. puertas_interiores	123. closet_material	124. cornisa	

Fuente: Elaboración propia del autor

### Organización de la base de datos:

**Base de datos general:** Se realizó la primera depuración a la base de datos donde se eliminaron 13 de las 34 variables iniciales debido a la baja relación e independencia entre ellas y el precio de los proyectos construcción, potencialmente no impactan directamente en la predicción del precio de los proyectos de construcción, las variables eliminadas se mencionan en la siguiente tabla:

Tabla 6 Eliminación de variables - Base de datos general – Depuración N° 1

1. Longitud	2. Latitud	3. Código Del Proyecto	4. Regional
5. Destino Etapa	6. Nombre Del Proyecto	7. Dirección	8. Nombre De La Etapa
9. Fecha terminación construcción	10. Fecha de entrega	11. ID tipo	12. Estado
13. Fase			

Fuente: Elaboración propia del autor.

Adicional se incluyeron las siguientes variables:

- SetTimeData: En esta variable se coloca la fecha 01/01/2020.
- DiferenciaVentas: Diferencia entre SetTimeData y fecha inicio de ventas con el fin de obtener el número de días que han transcurrido desde la fecha de inicio ventas hasta el 1 de enero de 2020.
- Anventas: Se obtiene al dividir el número de días que han transcurrido desde la fecha de inicio ventas hasta el 1 de enero de 2020 sobre 365 días, con el fin de obtener el número de años que han transcurrido.

Para la base de datos general se identificaron los valores nulos en las variables alcobas y baños con el objetivo de reemplazarlos por 0 (cero) y con la palabra “Sin definir” para las variables categóricas. En la siguiente tabla se presenta el nivel de poblamiento de datos ausentes:

Tabla 7 Número de registros ausentes

Uso	Número de registros ausentes			
	Baños		Alcobas	
	# Registros	%	# Registros	%
Residencial	462	0,89%	431	0,83%

No residencial	8.402	16,23%	8.905	17,20%
----------------	-------	--------	-------	--------

Fuente: Elaboración propia del autor.

Una vez realizados los ajustes anteriormente mencionados en la fase exploratoria, se procede a cargar la base general y la base de acabados en RStudio y continuar con la depuración y organización de la misma.

En Rstudio quedó cargada la base de datos general actualizada con la primera depuración la cual cuenta con una cantidad de 52.085 registros y 26 variables. La segunda etapa de depuración se fundamentó en la extracción de proyectos que presentan valor nulo o cero en el precio ya que se generan sesgos en el ajuste de los modelos y no aportan al pronóstico quedando un total de 40.537 registros y 24 variables. Posterior a eso se procede a clasificar las variables categóricas y numéricas quedando de la siguiente forma:

Tabla 8 Clasificación variables categóricas y numéricas - Base de datos general

VARIABLES CATEGÓRICAS			
Ciudad	Zona	Estrato	Barrio
Sistema Constructivo	Tipo por etapa	Usos	Segmento
Localidad			
VARIABLES NUMÉRICAS			
ID Proyecto	ID etapa	Numero de unid por etapa	Fecha inicio ventas
Fecha inicio construcción	Unid por tipo	Área por tipo	Área total
Alcobas	Baños	Precio de Venta	Precio por metro
Set Time Data	Diferencia Ventas	An ventas	

Fuente: Elaboración propia del autor

En la tercera depuración se eliminaron las siguientes variables puesto que su información no se requería en el modelo quedando una base de 40.537 registros y 17 variables (ver tabla 8).

Tabla 9 Eliminación de variables - Base de datos general – Depuración N° 3

VARIABLES ELIMINADAS		
ID Etapa	Fecha inicio Ventas	Fecha inicio construcción
Área total	Precio por metro	SetTimeData
Diferenciaventas		

Fuente: Elaboración propia del autor

La variable ID etapa se eliminó puesto que es una numeración propia de Camacol para clasificar los proyectos; Fecha inicio ventas, Fecha inicio de construcción, Diferencia ventas no se requieren puesto que solo se utilizara la variable Anventas que corresponde al número de años que han transcurrido desde la fecha inicio de ventas hasta el 1 de enero de 2020. La variable precio por metro se elimina debido a que está directamente relacionada con el precio de venta y la variable área total no se requiere puesto que está directamente relacionada con las variables área por tipo y número de unidades por tipo, la anterior decisión se toma para evitar correlación entre variables debido a que aportan la misma información.

A continuación se muestran los niveles que presenta la base de datos general para cada una de las variables categóricas:

Tabla 10 Niveles (Variables categóricas Ciudad y Zona) - Base de datos General

	<b>Ciudad</b>	<b>Total registros</b>
1	BOGOTÁ, D.C.	28.439
2	SOACHA	2.124
3	CHÍA	1.390
4	CAJICÁ	1.254
5	FUSAGASUGÁ	1.015
6	ZIPAQUIRÁ	763
7	TOCANCIPÁ	723
8	MOSQUERA	692
9	MADRID	691
10	COTA	437

	<b>Zona</b>	<b>Total registros</b>
1	10 - MULTICENTRO	3.694
2	05 - CEDRITOS	2.395
3	44 - SOACHA	2.124
4	12 - CHICO	2.119
5	02 - MAZUREN	1.667
6	36 - CORDOBA	1.627
7	54 - CHIA	1.390
8	49 - CAJICA	1.254
9	25 - AMERICAS	1.210
10	01 - MARANTA	1.109

Ver Anexo 1 - Fuente: Elaboración propia del autor

Tabla 11 Niveles (Variables categóricas Estrato y Barrio) - Base de datos General

	<b>Estrato</b>	<b>Total registros</b>
1	0	3.405
2	1	20
3	2	3.151
4	3	11.765
5	4	11.020
6	5	6.181
7	6	4.933

	<b>Barrio</b>	<b>Total registros</b>
1	20 DE JULIO	1
2	7 DE AGOSTO	3
3	ABRAHAN LINCON	4
4	ACACIAS USAQUEN	220
5	ACEVEDO TEJADA	9
6	AEROPUERTO EL DORADO	1
7	ALAMEDA	49
8	ALAMOS	45
9	ALCALA	4
10	ALCALA SPRING NORTE	5

Ver Anexo 2 - Fuente: Elaboración propia del autor

Tabla 12 Niveles (Variable categórica Sistema Constructivo) - Base de datos General

	<b>Sistema Constructivo</b>	<b>Total Registros</b>
1	Estructura combinada - Concreto con metal	12
2	Estructura combinada (Concreto con metal) - (Estructura combinada)	12
3	Estructura metálica	490
4	Industrializado - Forsa	2.255
5	Industrializado - Manoportable	487
6	Industrializado - Otro tipo de formaleta metálica	218
7	Industrializado - Outinord	447
8	Industrializado - Tunel	148
9	Industrializado - Unispan	129
10	Industrializado - Contech	578
11	Mampostería confinada	35

12	Mampostería estructural	4.269
13	No responde	13
14	Otro	127
15	Sin definir	7.713
16	Tradicional aporticado	23.542

Tabla 13 Niveles (Variables categóricas Tipo por etapa, Usos y Segmento) - Base de datos

	<b>Tipo por etapa</b>	<b>Total registros</b>
1	Apartamento	33.011
2	Bodega	253
3	Casa	2.916
4	Comercial	1
5	Consultorio	202
6	Habitaciones	19
7	Local	1.723
8	Lote	630
9	Oficina	1.708
10	Otro	1
11	Parqueadero	7
12	Servicios	2
13	Sin definir R	2

	<b>Usos</b>	<b>Total registros</b>
1	No residencial	4.548
2	Vivienda	35.927

	<b>Segmento</b>	<b>Total registros</b>
1	No Residencial	4.548
2	No VIS	25.117
3	SIN ASIGNAR	3.522
4	VIP	322
5	VIS	6.966

Para la cuarta depuración se realizó una subdivisión de la base de datos general clasificándola según su uso “residencial” y “no residencial”, lo anterior con el fin de realizar una imputación de datos con la mediana para la variable baño y alcobas que tengan valores 0 (cero) puesto que se consideró como inconsistencia que un proyecto residencial no tenga baño ni alcobas. Una vez clasificado se tienen 4.574 registros para el uso no residencial y 35.963 datos para uso residencial con 17 variables.

Se construyó en Excel una tabla donde se recopiló el valor de la inflación desde el año 1946 hasta el 2019 tal como se muestra en la siguiente tabla donde en la columna “prom” se entrega el valor promedio desde el año que indica la fila hasta el año 2019, con el objetivo de cargarla en R Studio.

Tabla 14 Valores históricos Inflación 1946 – 2019 – Fuente DANE

	X	Posición	n. relativa	AÑ.o	Inflación	n. anual	prom
1	1			39	1946	9.30	13.500270
2	2			50	1947	18.30	13.557808
3	3			44	1948	16.00	13.491944
4	4			20	1949	6.70	13.456620
5	5			54	1950	20.50	13.553143
6	6			36	1951	8.90	13.452464
7	7			1	1952	-2.30	13.519412
8	8			27	1953	7.30	13.755522
9	9			35	1954	8.80	13.853333
10	10			2	1955	1.72	13.931077
11	11			33	1956	8.47	14.121875
12	12			52	1957	18.75	14.211587
13	13			37	1958	9.21	14.138387
14	14			26	1959	7.23	14.219180
15	15			21	1960	6.74	14.335667
16	16			17	1961	6.32	14.464407
17	17			24	1962	6.93	14.604828
18	18			74	1963	32.41	14.739474
19	19			32	1964	8.39	14.423929
20	20			43	1965	14.84	14.533636
21	21			40	1966	12.92	14.527963
22	22			28	1967	7.46	14.558302

(...)

48 48	59 1993	22.61	8.464074
49 49	58 1994	22.60	7.920000
50 50	53 1995	19.47	7.332800
51 51	56 1996	21.64	6.827083
52 52	47 1997	17.68	6.183043
53 53	46 1998	16.70	5.660455
54 54	38 1999	9.23	5.134762
55 55	34 2000	8.75	4.930000
56 56	29 2001	7.65	4.728947
57 57	25 2002	6.99	4.566667
58 58	19 2003	6.49	4.424118
59 59	14 2004	5.50	4.295000
60 60	13 2005	4.85	4.214667
61 61	12 2006	4.48	4.169286
62 62	15 2007	5.69	4.145385
63 63	30 2008	7.67	4.016667
64 64	4 2009	2.00	3.684545
65 65	6 2010	3.17	3.853000
66 66	9 2011	3.73	3.928889
67 67	5 2012	2.44	3.953750
68 68	3 2013	1.94	4.170000
69 69	8 2014	3.66	4.541667
70 70	22 2015	6.77	4.718000
71 71	16 2016	5.75	4.205000
72 72	11 2017	4.09	3.690000
73 73	7 2018	3.18	3.490000
74 74	10 2019	3.80	3.800000

~ |

Posterior a esto, se crea una columna adicional a la base de datos general llamada Precio P y se alimenta con las inflaciones promedio con el objetivo de traer los precios de venta a valor presente. Debido a que no es recomendable realizar imputación de datos para los valores nulos en la variable precios de venta se decidió eliminar dichos valores quedando una base de 40.475 registros y 18 variables.

Base de datos acabados: Se carga en RStudio la base de acabados en formato .csv la cual tiene 21.190 registros con 124 variables, con la librería dplyr se eliminó todos los valores duplicados que tenía la base quedando 10.401 registros con 124 variables.

Con la función inner join se unificó las dos bases de datos general y acabado por la variable que tienen en común “ID proyecto” quedando una base de datos de 40.475 registros con 141 variables. Al unificar las dos bases de datos se identificó que existían variables que se duplicaban como ciudad y sistema constructivo razón por la cual se eliminaron. Para la quinta se eliminan las siguientes variables:

- Precio de venta: Esta variable ya no se requiere puesto que con estos valores se

construyó una nueva variable denominada “Precio P” la cual contiene todos los precios de venta traídos a valor presente (Enero 1 de 2020), estos nuevos valores son los que serán tomados como variable de salida para alimentar el modelo, con lo anterior, se busca no tener sesgos en cuanto a la pérdida del valor del dinero en el tiempo.

- An ventas: Esta variable fue utilizada con el fin de conocer el tiempo transcurrido en años desde la fecha de finalización de ventas de cada proyecto a la fecha en la que se desea traer los precios de venta a valor presente (Enero 1 de 2020); debido a que en esta etapa ya se conocen los Precios de Venta traídos a valor presente los cuales se encuentran consignados en la variable Precio P esta variable ya no es requerida.
- Activo/Inactivo: Esta variable no se requiere puesto que es una nomenclatura interna de Camacol para conocer si un proyecto se encuentra activo o inactivo.
- Código de Proyecto: Esta variable se decide eliminar puesto que es un código interno que maneja Camacol para registrar cada proyecto.
- Nombre del proyecto: Debido a que el objetivo es la predicción del precio de venta se decide eliminar el nombre del proyecto debido a que la información que proporciona esta variable no varía el precio de cada registro.
- Regional: Esta variable decide ser eliminada puesto que para todos los registros la información es la misma “Cundinamarca”.
- Ciudad y: La variable “ciudad” se encontraba en las dos bases de datos “general” y “acabados”, una vez se unificaron las bases de datos la información de esta variable quedó duplicada razón por la cual se decide eliminar una de ellas.
- Dirección del proyecto: De acuerdo al análisis realizado se establece que la ubicación del proyecto impacta en los precios de venta de cada registro, por lo anterior, se establece que para el modelo se dejarán variables tales como: Ciudad, zona y Barrio debido a que estas permiten realizar una mejor clasificación, razón por la cual se decide eliminar la variable “Dirección del proyecto”.
- Teléfono: Se define que esta variable no tiene una relación directa que impacte en el precio de venta que es nuestra variable a predecir.
- Destino: Esta variable entrega información en cuanto al uso final de cada proyecto (venta, uso propio, arriendo etc), de acuerdo al análisis realizado se establece que



el uso que cada propietario le designe al inmueble, el precio de venta no sufre ninguna modificación razón por la cual se decide eliminar esta variable del modelo.

- Uso general del proyecto: Esta variable se encontraba en las dos bases de datos “general” y “acabado”, una vez se unificaron las bases de datos la información de esta variable quedó duplicada razón por la cual se decide eliminar una de ellas.
- ID Proyecto: Esta variable se decide eliminar puesto que es un código interno que maneja Camacol para registrar cada proyecto.
- Sistema constructivo y: Esta variable se encontraba en las dos bases de datos “general” y “acabado”, una vez se unificaron las bases de datos la información de esta variable quedó duplicada razón por la cual se decide eliminar una de ellas.
- Abierta/cerrada: Esta variable no se requiere puesto que es una nomenclatura interna de Camacol para conocer si un proyecto se encuentra abierto o cerrado.
- Código etapa: Esta variable se decide eliminar puesto que es un código interno que maneja Camacol para registrar cada proyecto.
- Nombre etapa: Debido a que el objetivo es la predicción del precio de venta se decide eliminar el nombre de la etapa debido a que la información que proporciona esta variable no varía el precio de cada registro.
- Unidades etapa, destino etapa, uso etapa: Estas variables se encontraban en las dos bases de datos “general” y “acabado”, una vez se unificaron las bases de datos la información de esta variable quedó duplicada razón por la cual se decide eliminar una de ellas.

Una vez terminada la depuración y unificación de las bases de datos queda una base de 40.475 datos con 117 variables.

### Identificación de las variables de entrada y salida que se van a utilizar en cada modelo

Tabla 15 Variables de entrada y salida definitivas para los modelos

	VARIABLE	TIPO		VARIABLE	TIPO
1	PrecioP	Variable salida	31	Zona oficina administración	Variable entrada
2	Ciudad	Variable entrada	32	Zona espejo de agua	Variable entrada
3	Zona	Variable entrada	33	Zona parque infantil	Variable entrada
4	Estrato	Variable entrada	34	Zona citofonia	Variable entrada
5	Barrio	Variable entrada	35	Zona aire acondicionado	Variable entrada
6	Sistema Constructivo	Variable entrada	36	Zona cinema	Variable entrada
7	Tipo por etapa	Variable entrada	37	Zona sendero de trote	Variable entrada

8	Usos	Variable entrada	38	Zona terraza sky club	Variable entrada
9	Segmento	Variable entrada	39	Zona fuente	Variable entrada
10	Localidad	Variable entrada	40	Zona red contra incendios	Variable entrada
11	Número de unidades por etapa	Variable entrada	41	Zona de pesca	Variable entrada
12	Unidades por tipo	Variable entrada	42	Zona kiosko	Variable entrada
13	Área por tipo	Variable entrada	43	Zona cuarto de basuras	Variable entrada
14	Alcobas	Variable entrada	44	Zona cancha de tenis	Variable entrada
15	Baños	Variable entrada	45	Zona shut basuras	Variable entrada
16	Zona internet	Variable entrada	46	Zona tanque abastecimiento de agua	Variable entrada
17	Zona bbq	Variable entrada	47	Zona salón comunal	Variable entrada
18	Zona balcón	Variable entrada	48	Zona caldera	Variable entrada
19	Zona sede social	Variable entrada	49	Zona golfito	Variable entrada
20	Zona ascensores	Variable entrada	50	Zona portería	Variable entrada
21	Zona gimnasio	Variable entrada	51	Zona salón de juegos	Variable entrada
22	Zonas verdes	Variable entrada	52	Zona campo de golf	Variable entrada
23	Zona guardería	Variable entrada	53	Zona lavandería	Variable entrada
24	Zona ciclo ruta	Variable entrada	54	Zona ciclero	Variable entrada
25	Zona piscinas	Variable entrada	55	Zonas húmedas	Variable entrada
26	Zona bar	Variable entrada	56	Zona salones especiales sala de juntas	Variable entrada
27	Zona cancha múltiple	Variable entrada	57	Zona planta eléctrica de emergencia	Variable entrada
28	Zona cancha squash	Variable entrada	58	Zona plazoletas	Variable entrada
29	Zona cancha de futbol	Variable entrada	59	Zona centro computo	Variable entrada
30	Zona lobby recepción	Variable entrada	60	Zona cafetería	Variable entrada
	<b>VARIABLE</b>	<b>TIPO</b>		<b>VARIABLE</b>	<b>TIPO</b>
61	Zona restaurante	Variable entrada	91	Subestructura sótano	Variable entrada
62	Zona auditorios	Variable entrada	92	Divisiones inferiores	Variable entrada
63	Zona capilla	Variable entrada	93	Fachada	Variable entrada
64	Zona baterías de baños	Variable entrada	94	Cubiertas	Variable entrada
65	Zona vestier	Variable entrada	95	Ventanería	Variable entrada
66	Zona biblioteca	Variable entrada	96	Pisos alcobas	Variable entrada
67	Zonas cuarto gases	Variable entrada	97	Muros alcobas	Variable entrada
68	Zona mascotas	Variable entrada	98	Cielo raso alcobas	Variable entrada
69	Zona cuarto de equipos	Variable entrada	99	Pisos cocina	Variable entrada
70	Zona de niños	Variable entrada	100	Muros cocina	Variable entrada
71	Zona área lounge	Variable entrada	101	Cielo raso cocina	Variable entrada
72	Zona sala de estudios	Variable entrada	102	Tipo cocina	Variable entrada
73	Zona escaleras	Variable entrada	103	Mesón cocina	Variable entrada
74	Zona minimarket	Variable entrada	104	Muebles cocina	Variable entrada
75	Zona conjunto cerrado	Variable entrada	105	Dotación cocina	Variable entrada
76	Zona sala múltiple	Variable entrada	106	Pisos baños	Variable entrada
77	Zona teatrino	Variable entrada	107	Muros baños	Variable entrada
78	Zona coworking	Variable entrada	108	Cielo raso baños	Variable entrada
79	Zona yoga	Variable entrada	109	Mesón baños	Variable entrada
80	Zona spa	Variable entrada	110	División baños	Variable entrada
81	Zona sauna	Variable entrada	111	Lavamanos	Variable entrada
82	Zona deposito	Variable entrada	112	Mueble lavamanos	Variable entrada
83	Zona vía de acceso	Variable entrada	113	Lavadero	Variable entrada
84	Zona tanque reserva de agua	Variable entrada	114	Puerta principal	Variable entrada
85	Zona roof garden	Variable entrada	115	Puertas inferiores	Variable entrada
86	Zonana	Variable entrada	116	Closet material	Variable entrada
87	Zona muro escalar	Variable entrada	117	Cornisa	Variable entrada
88	Zona no responde	Variable entrada			
89	Placa entrepiso	Variable entrada			
90	Subestructura cimentación	Variable entrada			

Fuente: Elaboración propia del autor

## Modelos supervisados

Con el fin de explicar el precio de los proyectos de construcción a partir de las diversas co-variables expuestas en la investigación se ajustan cuatro tipos de modelos, iniciando por el modelo de regresión lineal múltiple, los árboles de regresión, las máquinas de soporte vectorial y las redes neuronales.

### Ajuste del modelo de regresión lineal múltiple

El modelo busca la explicación del precio a partir de la combinación lineal de las variables independientes observadas más un componente del error que se asumen sigue una distribución normal con varianza de los errores constantes “homocedasticidad”.

Modelo.

$$y_i = \alpha + \beta_1 x_{ciudad} + \beta_2 x_{zona} + \beta_3 x_{estrato} + \beta_4 x_{barrio} + \beta_5 x_{SistemaConst.} + \beta_6 x_{tipoEtapa} + \beta_7 x_{usos} + \beta_8 x_{segmento} + \beta_9 x_{localidad} + \beta_{10} x_{numero\_unidades\_etapa} + \beta_{11} x_{area\_tipo} + \beta_{12} x_{unidades\_tipo} + \beta_{13} x_{alcobas} + \beta_{14} x_{baño} + \sum \beta_{zona}^k x_k + \dots + \varepsilon_i$$

Se realizan tres ajustes del modelo, uno tomando todos los registros, otro con registros de proyectos de uso residencial y otro con los proyectos de uso no residencial obteniendo los siguientes resultados de los ajustes:

Estadístico	Modelo completo	Uso residencial	Uno NO residencial
F	6,282	141,200	1,866
p.valor	< 2.2e-1	< 2.2e-1	< 2.2e-16
R <sup>2</sup>	0,163	0,824	0,097
R <sup>2</sup> ajustado	0,137	0,818	0,208
Error estándar de residuales	2.966.000	227.200	8.892.000

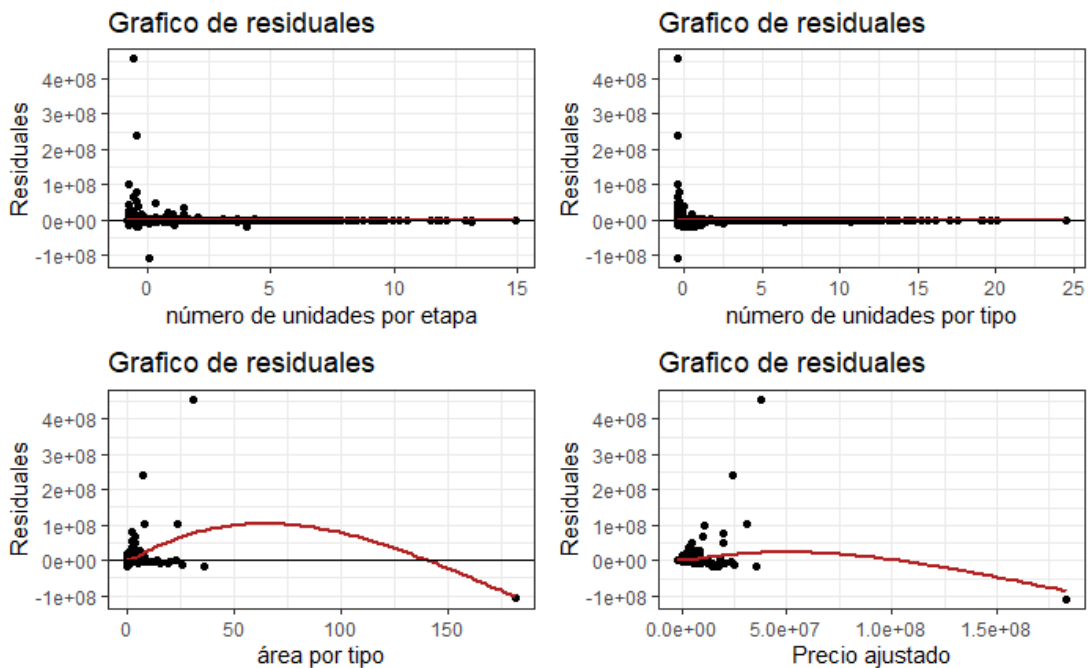
El ajuste del modelo requiere de otras verificaciones, no obstante, se detallara los supuestos del error donde recae las estrategias de estimación.

Al observar algunas pruebas de normalidad sobre los residuales de la regresión se evidencia p.valores muy bajos, casi cero, lo que indica el rechazo de la normalidad, al observar algunos gráficos puede deberse a algunos valores que son extremos pero pueden

ser consistentes con la realidad por lo que se recomienda otra estrategia para la predicción.

Prueba de normalidad	Modelo completo		Uso residencial		Uso No residencial	
	Estadístico	p.valor	Estadístico	p.valor	Estadístico	p.valor
Lilliefors	0,37982	< 2.2e-16	0,1689	< 2.2e-16	0,30686	< 2.2e-16
Anderson-Darling	11.380	< 2.2e-16	2.650,3	< 2.2e-16	951.35	< 2.2e-16
Normalidad de Pearson	275.000	< 2.2e-16	22.989	< 2.2e-16	17.805	< 2.2e-16
Jarque Bera	4,23E+15	< 2.2e-16	77.217.612	< 2.2e-16	458.518.477	< 2.2e-16
Frosini	42,637	< 2.2e-16	19,614	< 2.2e-16	12.312	< 2.2e-16

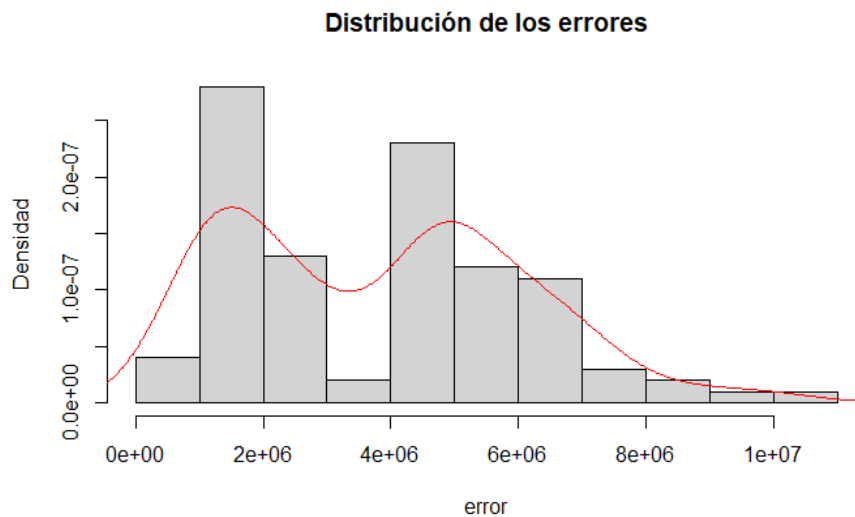
Con el fin de validar la homocedasticidad se observan los residuales frente a algunas co-variables principales y la relación entre los residuales y la variable repuesta los cuales se muestran a continuación:



El estadístico F del ajuste del modelo es de 6.282 con un p.valor < 2.2e-16 lo que indica el rechazo de la hipótesis que ninguna variable aporte al precio, es decir por lo menos hay alguna que influencia en el precio. No obstante, el R y el R cuadrados ajustado son de 0.1629 y 0.137, lo cual son muy bajos, adicionalmente al realizar un chequeo cruzado tomando un conjunto de entrenamiento y calculando el error de la estimación con los datos observados en el conjunto de prueba, para mirar la consistencia se seleccionan diferentes

conjuntos de entrenamiento entre el 60% y 80% de los datos con los que se ajusta el modelo de regresión y se calculan la raíz del promedio de los errores cuadráticos medio para cada uno de ellos. La siguiente tabla muestra el resumen estadístico de los errores junto al resumen de los precios de los datos observando una alta dispersión y errores bastante altos con relación a la respuesta original, no generando un buen ajuste.

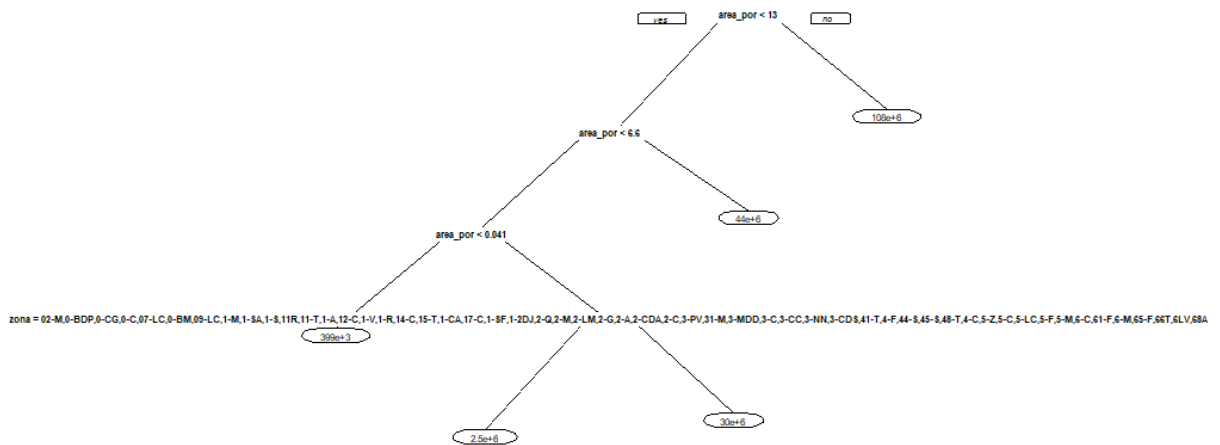
Estadístico	Errores generados	Precios
Min	\$ 922.396	\$ 29
Q1	\$ 1.514.702	\$ 153.017
Mediana	\$ 4.223.718	\$ 306.815
Q3	\$ 5.221.828	\$ 551.638
Max	\$ 10.054.493	\$ 492.778.685
Promedio	\$ 3.829.751	\$ 551.387



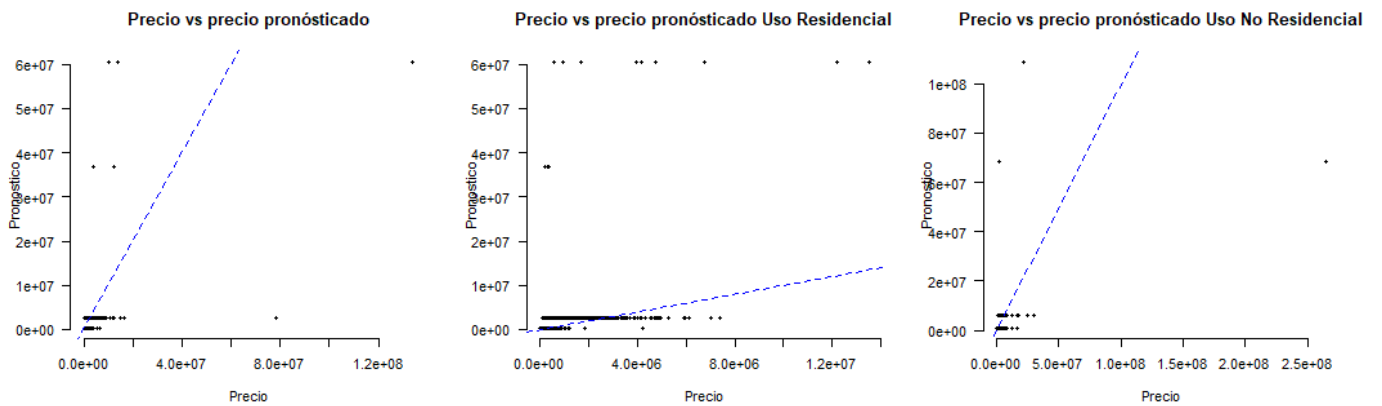
### Ajuste del modelo de árboles de decisión

Los árboles de regresión permiten explicar la respuesta de una variable a partir de responder preguntas dicotómicas a cerca de cada una de las co-variables. Al final construye hiper-rectangulos de tal forma que las observaciones que puedan estar dentro de un hiper-rectangulo tengan el mismo valor estimado (Hernández, 2021)

El siguiente diagrama muestra el árbol que se construye con las variables presentadas, en este caso son significantes las variables de la “zona” y el “área\_por\_tipo”.



Se ajusta un árbol de regresión con el 75% y se usa el 25% restante para probar el modelo del conjunto total de datos, uso residencial y no residencial. Al obtener el pronóstico de los precios bajo este modelo frente a los valores reales tenemos una correlación del 0.35, 0.47 y 0,519 respectivamente. Los correspondientes pronósticos sobre los conjuntos de prueba como se muestra en el siguiente gráfico.

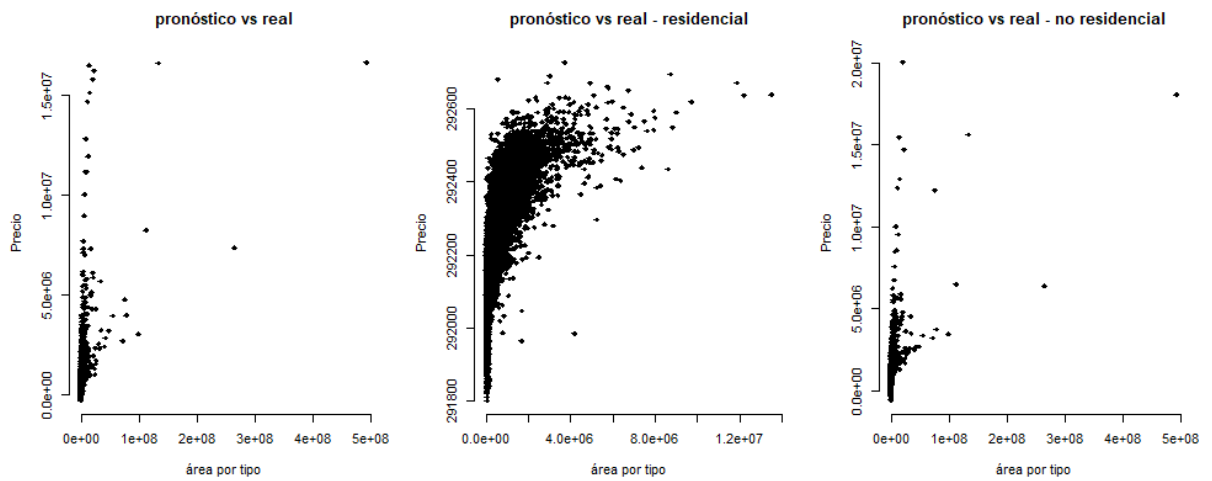


### Ajuste por máquina de soporte vectorial

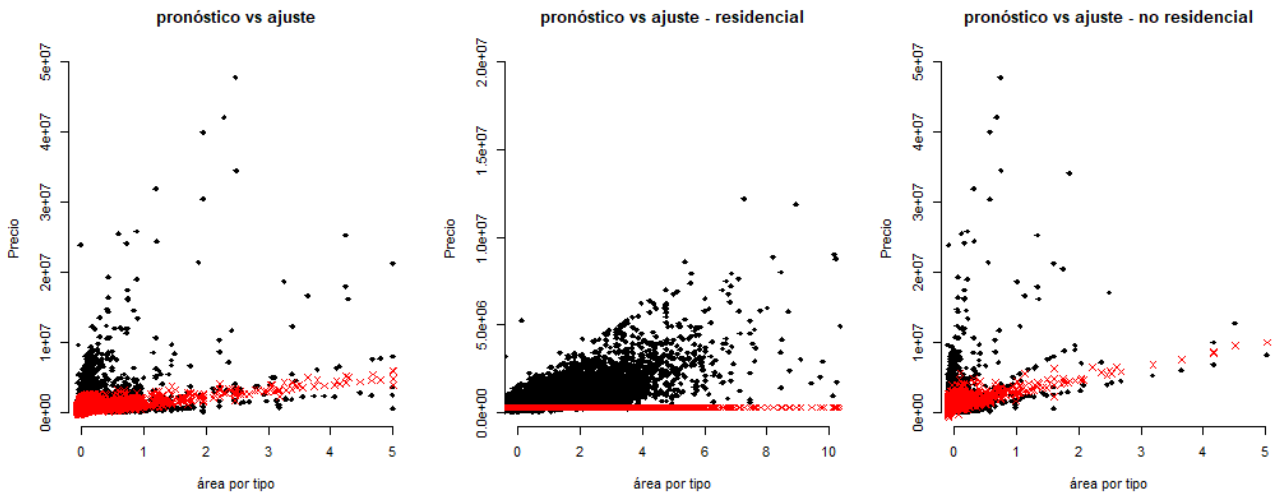
Con el fin de pronosticar el precio se ajusta otro modelo mediante una regresión de soporte vectorial (SVM) teniendo en cuenta las co-variables anteriormente mencionadas en los diversos conjuntos de datos, obteniendo los siguientes resúmenes de los modelos ajustados:

Parámetros	Modelo general	Modelo Uso Residencial	Modelo Uso No Residencial
gamma	0,00080	0,00084	0,00156
epsilon	0,1	0,1	0,1
Numero de vectores de soporte	5.293	35.926	926

Una vez se ajustan los modelos con los diversos conjuntos de datos se contrasta los pronósticos sobre los datos originales con los datos reales.



Los errores de pronóstico en los modelos ajustados son de 3.025.690, 550.616,3 y 8.960.913, y en la siguiente grafica podemos ver los valores del pronóstico (puntos color negro) frente a una de las variables de interés y el valor real del precio (puntos color rojo).



## **Pronóstico mediante redes neuronales**

Una red neuronal artificial es un sistema de computación que aprende de los datos administrados.

Primero, se genera una estructura de “neuronas” artificiales que se conectan entre sí. Luego, se pide que este conjunto resuelva un problema mediante intentos sucesivos fortaleciendo cada vez las conexiones que conducen al éxito. Tensorflow es un paquete que se puede instalar en Python para el ajuste de modelos de aprendizaje automático principalmente en el diseño y aprendizaje de redes neuronales. Ha sido desarrollado por Google Brain Team dentro de la organización de investigación de Machine Intelligence de Google para el aprendizaje automático y la investigación de redes neuronales. Tensorflow puede entrenar y ejecutar redes neuronales profundas para la clasificación de dígitos manuscritos, reconocimiento de imágenes, incrustaciones recurrentes, modelo de secuencia a secuencia para traducción automática, procesamiento de lenguaje natural y simulaciones basadas en PDE (Ecuación Diferencial Parcial) (Sánchez & Israel, 2020).

### **Red Neuronal para los datos de Uso residencial**

Para la elaboración de este modelo se realizó una red neuronal artificial, tomando como información de entrada la base de datos exportada del modelo 2 en Rstudio en la cual se encuentran los proyectos de uso residencial (apartamentos y casas) que corresponden a 35.927 registros con 116 variables de entrada y una variable de salida (precio de venta) las cuales se relacionan en la tabla 15.

Se utiliza la librería `get.dummies` con el objetivo que cada variable categórica se subdivida en variables dummies quedando una base de 37.927 registros y 1.227 columnas o variables. La variable respuesta corresponde al precio de venta y las variables restantes se dejan como variables que ingresan a la red. Se realiza una subdivisión de la base de datos en la cual se deja el 80% de la base de datos para entrenamiento (data train) y el 20% restante para prueba o data test.

Los parámetros se muestran a continuación



```

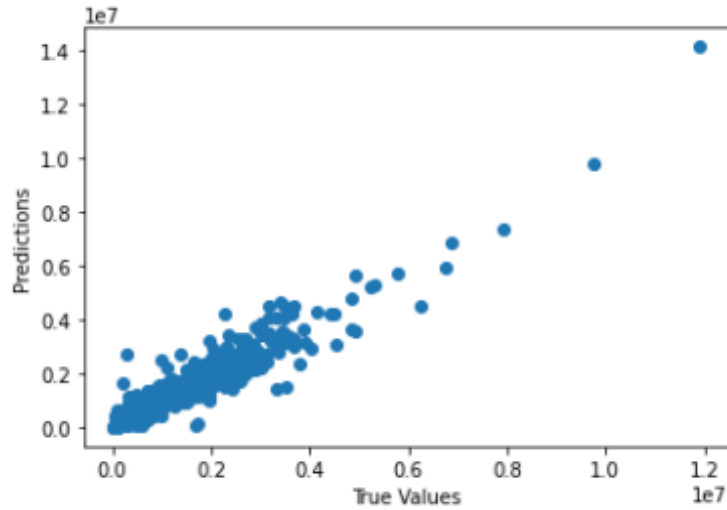
Model: "sequential"
-----
Layer (type)                Output Shape                Param #
-----
dense_1 (Dense)             (None, 1024)                1256448
dense_2 (Dense)             (None, 512)                 524800
dense_3 (Dense)             (None, 256)                 131328
dense_4 (Dense)             (None, 128)                 32896
dense_5 (Dense)             (None, 16)                  2064
dense_output (Dense)        (None, 1)                   17
-----
Total params: 1,947,553
Trainable params: 1,947,553
Non-trainable params: 0
-----

```

Una vez contruidos los errores a partir de la variación porcentual entre los valores pronosticados y los valores reales se observa que el error máximo obtenido es de 59,54 % de subestimación y 33.62 % de sobreestimación. Al tener en cuenta las 20 sub-muestras, el 5% de los individuos que más tienen pronostico por debajo del valor real en el conjunto de entrenamiento, se alejan en promedio en un 38,8% del valor real; por otro lado, el 5% de los individuos que tienen un pronóstico superior al valor real en el conjunto de prueba supera en un 25% al precio real.

Los correspondientes pronósticos sobre los conjuntos de prueba se muestran en el siguiente gráfico:

Precio real vs Precio pronosticado uso residencial



El 96,11% de los individuos del conjunto de prueba tienen el valor real del precio dentro del intervalo del pronóstico correspondiente.

Por otro lado, la red de uso No residencial hace uso la librería `get.dummies` con el objetivo que cada variable categórica se subdivida en variables dummies quedando una base de 4548 registros y 673 columnas o variables. Se escoge la variable “y” o variable a predecir que en este caso sería el precio de venta y las variables restantes se dejan como variables independientes. Se realiza una subdivisión de la base de datos en la cual se deja el 80% de la base de datos para entrenamiento (data train) y el 20% restante para prueba o data test.

Los parámetros se muestran a continuación:

Model: "sequential"

Layer (type)	Output Shape	Param #
dense_1 (Dense)	(None, 1024)	689152
dense_2 (Dense)	(None, 512)	524800
dense_3 (Dense)	(None, 256)	131328
dense_4 (Dense)	(None, 128)	32896
dense_5 (Dense)	(None, 16)	2064
dense_output (Dense)	(None, 1)	17

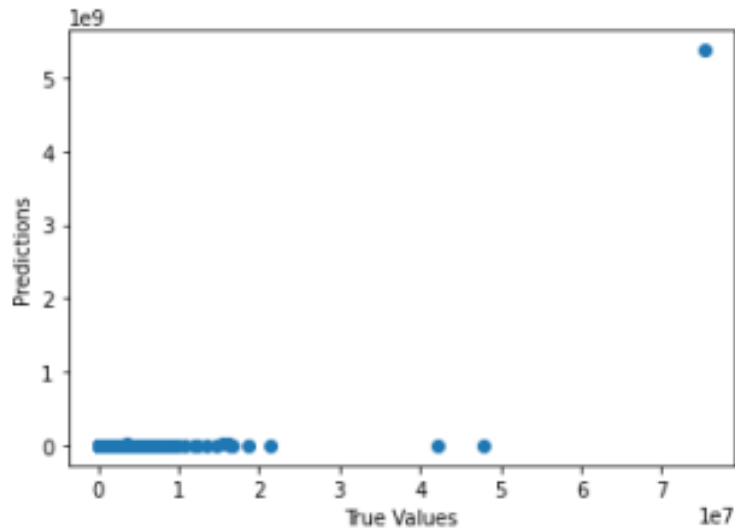
-----

Total params: 1,380,257  
Trainable params: 1,380,257  
Non-trainable params: 0

Una vez contruidos los errores a partir de la variación porcentual entre los valores pronosticados y los valores reales se observa que el error máximo obtenido es de 36,28 % de subestimación y 79,8 % de sobreestimación.

Los correspondientes pronósticos sobre los conjuntos de prueba se muestran en el siguiente gráfico:

Precio real vs Precio pronosticado uso no residencial



El 57,47 % de los individuos del conjunto de prueba tienen el valor real del precio dentro del intervalo del pronóstico correspondiente.

## **6. Conclusiones y recomendaciones**

La construcción es uno de los sectores que desempeñan un papel fundamental en la economía de cualquier país, sin embargo, se evidencia que la inversión e implementación en investigación y desarrollo es baja comparada con otras industrias, razón por la cual un gran porcentaje de los procesos desde la planeación hasta la ejecución aún se realiza sin ningún tipo de automatización generando incertidumbre y riesgos para el gremio constructor.

La presente investigación aportó en la fase de oferta económica dentro el proceso de negociación de los proyectos de construcción mediante el pronóstico de los precios a partir del uso de métodos supervisados usando 116 variables de entrada (ver tabla 15) que típicamente inciden en el mismo. El aporte se debe a la complejidad en esta fase, ya que durante la planeación existe una alta incertidumbre por parte de los profesionales que realizan las evaluaciones en etapas de pre-factibilidad puesto que no se cuenta con suficiente información técnica ni herramientas automatizadas que les permita realizar estimaciones asertivas evitando posibles desviaciones que se puedan ver representadas en pérdidas económicas para los inversionistas y gerentes.

Para el pronóstico se ajustaron modelos supervisados en Machine Learning usando una base de datos de 51.579 registros que proporcionó Camacol obtenida del Censo Nacional de Edificaciones que se realiza desde el año 1995.

Las redes neuronales ajustaron de una mejor forma el precio, aportando notoriamente a la automatización de los procesos de planeación de los proyectos en el sector de la construcción contribuyendo a la minimización de errores generados por procesos manuales, sin embargo, se requiere tener bases de datos consolidadas, robustas y de calidad que le permitan a la academia realizar modelos cada vez más precisos. Lo anterior se evidencia frente a la alta dificultad que se encontró durante la elaboración de esta tesis en cuanto a la consecución de bases de datos confiables que permitieran alimentar los modelos.

Para la predicción por redes neuronales se estructuraron dos modelos, uno de ellos toma en cuenta los proyectos de uso residencial y el otro para los proyectos de uso no residencial. La

arquitectura ajustada está conformada por 5 capas ocultas con 1.024, 512, 256, 128, y 16 neuronas respectivamente, finalizando con una capa de una neurona en los dos casos. El entrenamiento de la red tomo en cuenta el 80% de la información, donde se seleccionan 20 subconjuntos, en cada uno se toma el 80% para entrenamiento y 20% de prueba con el fin de tener consistencia en los resultados.

Para el primer modelo con redes neuronales, el realizado sobre los proyectos de uso residencial, la red logra ajustar bien el precio de los proyectos ya que cerca del 96,11 % de los individuos del conjunto de prueba tienen el valor real del precio dentro del intervalo del pronóstico correspondiente. Al tener en cuenta las 20 sub-muestras, el 5% de los individuos que más tienen pronostico por debajo del valor real en el conjunto de entrenamiento, se alejan en promedio en un 38,8% del valor real; por otro lado, el 5% de los individuos que tienen un pronóstico superior al valor real en el conjunto de prueba supera en un 25% al precio real.

Para el caso de los proyectos No residenciales, el pronóstico por redes neuronales no es tan adecuado ya que tan solo cerca del 57,47 % de los individuos del conjunto de prueba tienen el valor real del precio dentro del intervalo del pronóstico correspondiente y se encuentran errores que están por encima del 79,8 % del valor real. La razón puede obedecer al número de proyectos con los que fue entrenada la red la cuál es muy inferior al número de datos con los que se entrenó el modelo residencial, por lo anterior, se recomienda para tesis futuras realizar nuevos modelos con mayor cantidad de datos de proyectos de uso no residencial, esto podría ser posible puesto que Camacol cuenta con una base de datos más extensa a la analizada en la presente tesis puesto que realiza el Censo Nacional de edificaciones a nivel Nacional.

El primer modelo de regresión lineal que se realizó en el software RStudio para los proyectos de uso residencial y no residencial con una totalidad de 40.475 proyectos el cual arrojó un R cuadrado de 0,163, una vez se practicó el test de lilliefors se identificó que el p-value era inferior a 0.05 lo que indicó que sus errores no se distribuían normalmente. Adicional a lo anterior, se validó la prueba de hipótesis que indica que en el modelo se están usando variables independientes que si explican la variable de salida (Precio de venta).

Para el segundo y tercer modelo de regresión lineal se dividió la base de datos en proyectos de uso residencial (modelo 2) y uso no residencial (modelo 3). Para el modelo 2 se obtuvo un R-cuadrado de 0,824 y para el tercer modelo un R-Cuadrado de 0.097, por lo anterior se evidencia que el modelo de regresión lineal con mejor R-cuadrado es el número 2 que corresponde a proyectos de uso residencial, razón por la cual se infiere que los modelos tuvieron un mejor comportamiento al dividirlos de acuerdo a su uso, sin embargo el modelo número 3 no tuvo un buen R-cuadrado y esto pudo deberse a que el número de proyectos con el cual se realizó el modelo es muy pequeño (4.548 proyectos) en comparación al segundo modelo que se trabajó con 35.927 proyectos, por lo anterior, se recomienda a futuro realizar modelos para uso no residencial con mayor cantidad de datos con el fin de identificar si esta pudo haber sido la causa del comportamiento encontrado. Sin embargo una vez se realizó el test de lilliefors para los dos modelos se identificó que el p-value era menor a 0.05 lo que indicó que sus residuales no se distribuían normalmente. Por los resultados expuestos anteriormente se hizo necesario evaluar modelos con redes neuronales. Adicionalmente se ajustaron modelos supervisados como árboles de decisión y máquina de soporte vectorial con el fin de explorar alternativas que dependieran de la variabilidad de los datos, los cuales generaron errores bastante altos entre el valor real y el pronosticado, generando inconsistencias en las estimaciones para este fin. Los árboles de regresión para el modelo 1 (residencial + no residencial), modelo 2 residencial y modelo 3 no residencial tuvieron una correlación de 0.35, 0.47 y 0.51 respectivamente. Para los modelos de ajuste por máquina de soporte vectorial los errores de pronóstico son de 3.025.690 para el modelo 1, 550.616,3 modelo 2 y 8.960.913 modelo 3.

## **7. Referencias bibliográficas**

Akanmu, A., & Anumba, C. J. (2015). Cyber-physical systems integration of building information models and the physical construction. *Engineering, Construction and Architectural Management*, 22(5), 516–535. <https://doi.org/10.1108/ECAM-07-2014-0097>.

Artola Moreno, A. (2019). Clasificación de imágenes usando redes neuronales convolucionales en Python. (Tesis de posgrado). Universidad de Sevilla. Recuperado de:

<http://bibing.us.es/proyectos/abreproy/92402/fichero/TFG-2402-ARTOLA.pdf>

Brownlee, J. (2020). Machine Learning Mastery – Making Developers Awesome at MachineLearning. <https://machinelearningmastery.com/>

Cámara Nacional de la Construcción, Camacol. (s.f). Camacol. Recuperado de:<https://camacol.co/nosotros/quienes-somos>

Camacol. (2018). Informe de productividad. Sector construcción de edificaciones. Recuperado de:<https://camacol.co/sites/default/files/INFORME-PRODUCTIVIDAD-VF.PDF>

Camargo Sánchez, J.A. (2015). Análisis de problemas de construcción derivados de errores de diseño. (Tesis de pregrado). Universidad de los Andes. Recuperado de:  
<https://repositorio.uniandes.edu.co/bitstream/handle/1992/13493/u722589.pdf?sequence=1>

Colaboratory (s.f). Recuperado de: <https://www.rstudio.com/products/rstudio/>

Dallasega, P., Rauch, E., & Linder, C. (2018). Industry 4.0 as an enabler of proximity for construction supply chains: A systematic literature review. Computers in Industry, 99, 205–225. Recuperado de: <https://doi.org/https://doi.org/10.1016/j.compind.2018.03.039>

Galán Asensio, H. y Martínez Bowen, A. (2018). Inteligencia artificial.Redes neuronales y aplicaciones. Recuperado de: <http://www.it.uc3m.es/jvillena/irc/practicas/10-11/06mem.pdf>

Galindo Ramírez, N.M. (2016). Inteligencia artificial aplicada a las organizaciones. (Tesis de posgrado). Universidad Militar Nueva Granada. Recuperado de: <https://repository.unimilitar.edu.co/bitstream/handle/10654/15965/GalindoRam%EDrezNohraMay2016.pdf;jsessionid=368E98C41972932A9B88BA2D164DB618?sequence=1>

García Betancur, M.A.; Hincapié Gómez, V.O. y Pérez Ochoa, L.C. (2015). Metodología para la mitigación de los errores en los procesos de construcción de proyectos de edificaciones. (Tesis de posgrado). Universidad de Medellín. Recuperado de:

[https://repository.udem.edu.co/bitstream/handle/11407/2162/TG\\_EIOC\\_1.pdf?sequence=](https://repository.udem.edu.co/bitstream/handle/11407/2162/TG_EIOC_1.pdf?sequence=1)

[1](https://repository.udem.edu.co/bitstream/handle/11407/2162/TG_EIOC_1.pdf?sequence=1)

Gaviria Londoño, S. (2016). Costos para la construcción. (Tesis de maestría). Universidad EAFIT.

Recuperado

de:

[https://repository.eafit.edu.co/bitstream/handle/10784/9531/Sebastian\\_GaviriaLondo%C3%B1o\\_2016.pdf?sequence=2](https://repository.eafit.edu.co/bitstream/handle/10784/9531/Sebastian_GaviriaLondo%C3%B1o_2016.pdf?sequence=2)

Giraldo Hernández, C. (2020). Implementación computacional en python de desarrollos mecánico-cuánticos empleados en el curso de átomos y moléculas para estudiar el comportamiento de la materia a nivel atómico y molecular. Universidad Nacional de Colombia. [Recuperado de: https://repositorio.unal.edu.co/handle/unal/79670](https://repositorio.unal.edu.co/handle/unal/79670)

Godoy Viera, A.F. (2014). Técnicas de aprendizaje de máquina utilizadas para la minería de texto.

BIBLIOTECOLÓGICA, vol. 31, núm. 71. Recuperado de: <http://www.scielo.org.mx/pdf/ib/v31n71/2448-8321-ib-31-71-00103.pdf>

Groves-Delphos, S. (2014). Cloud Computing, BIM trends in construction industry, The BIM Hub (March) (2014). Guía de los fundamentos para la Dirección de Proyectos. (2008). (Guía del PMBOK) Project Management Institute.

Gutiérrez Gutiérrez, G. y Rodríguez Gutiérrez, A. (2018). Introducción Aprendizaje de Máquina. Recuperado de:

<http://c3.itm.edu.co/~gerardo.gutierrez/machinelearning/Introducci%C3%B3n%20Aprendizaje%20de%20M%C3%A1quina.pdf>

Hernández Sampieri, R.; Fernández Collado, C. y Baptista Lucio, P. (2016). Metodología de la investigación. Sexta edición. Mc Graw Hill. Recuperado de: <https://www.uca.ac.cr/wp-content/uploads/2017/10/Investigacion.pdf>

Hernández Barajas, F. (2021). Modelos predictivos. Recuperado de: ([https://fhernanb.github.io/libro\\_mod\\_pred/arb-de-regre.html](https://fhernanb.github.io/libro_mod_pred/arb-de-regre.html))

Inafuku, A. (2020). Diseño e implementación de un sistema de Diagnóstico de fallas para la



inspección y detección de fallas en componentes de procesos industriales utilizando un Robot Móvil y algoritmos de inteligencia artificial. (Tesis de maestría). Pontificia Universidad Católica del Perú, Escuela de Posgrado.

Invest in Bogotá. (2021). Materiales de construcción, un sector diverso con múltiples oportunidades de negocio. Recuperado de: <https://es.investinbogota.org/sectores-de-inversion/materiales-de-construccion>

King Chun, C., Li, H., & Skitmore, M. (2012). The use of virtual prototyping for hazard identification in the early design stage. Construction Innovation, 12(1), 29–42. <https://doi.org/10.1108/14714171211197481>

K. Jung, B. Chu, D. H. (2013). Robot-based construction automation: an application to steel beam assembly (part II).

López de Ortigosa Casares, D. (2010). Ingeniería de costos en la construcción n. México DF: Trillas

Lozano Serna, S.; Patiño Galindo, I.; Gómez Cabrera, A. y Torres, A. (2018). Identificación de factores que generan diferencias de tiempo y costos en proyectos de construcción en Colombia. *ing. cienc.*, vol. 14, no. 27, pp. 117–151. Recuperado de: <http://www.scielo.org.co/pdf/ince/v14n27/1794-9165-ince-14-27-00117.pdf>

Majrouhi Sardroud, J. (2012). Influence of RFID technology on automated management of construction materials and components. *Scientia Iranica*, 19(3), 381–392. <https://doi.org/10.1016/J.SCIENT.2012.02.023>

Matich, J. (2001). *Redes Neuronales: Conceptos Básicos y Aplicaciones*. (Tesis de posgrado). Universidad Tecnológica Nacional. Recuperado de: [https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5\\_ano/orientadora1/monograis/matich-redesneuronales.pdf](https://www.frro.utn.edu.ar/repositorio/catedras/quimica/5_ano/orientadora1/monograis/matich-redesneuronales.pdf)

Mendoza H, Vargas J, López L, Bautista G. (2002). *Métodos de Regresión*. Universidad Nacional de Colombia, <http://www.virtual.unal.edu.co/cursos/ciencias/2007315/>. Licencia: Creative Commons BY-NC-ND.

Merschbrock, C., & Munkvold, B. E. (2015). Effective digital collaboration in the construction

industry – A case study of BIM deployment in a hospital construction project. *Computers in Industry*, 73, 1–7. <https://doi.org/https://doi.org/10.1016/j.compind.2015.07.003>

Mitchell, T. M. *Machine Learning*. New York: McGraw Hill, 1997

Mora Salas, D. (2015). *Diseño de un plan de gestión de costo y de tiempo para la administración de proyectos de la empresa constructora Jiménez y Sojo*. (Tesis de pregrado). Instituto

Tecnológico de Costa Rica. Recuperado de:

[https://repositoriotec.tec.ac.cr/bitstream/handle/2238/6160/Dise%F1o\\_plan\\_gesti%F3n\\_costo%20tiempo\\_administraci%F3n\\_constructora%20Jim%29nez%20y%20Sojo.pdf?sequence=1](https://repositoriotec.tec.ac.cr/bitstream/handle/2238/6160/Dise%F1o_plan_gesti%F3n_costo%20tiempo_administraci%F3n_constructora%20Jim%29nez%20y%20Sojo.pdf?sequence=1)

OIT. (2018). *Mejore su negocio de construcción*. Recuperado de:

<https://www.ilo.org/public/spanish/employment/recon/eip/download/manual1.pdf>.

Oesterreich, T. D., & Teuteberg, F. (2016). Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry. *Computers in Industry*, 83, 121–139. <https://doi.org/10.1016/j.compind.2016.09.006>.

Park, C.-S., Lee, D.-Y., Kwon, O.-S., & Wang, X. (2013). A framework for proactive construction defect management using BIM, augmented reality and ontology-based data collection template. *Automation in Construction*, 33, 61–71.

<https://doi.org/https://doi.org/10.1016/j.autcon.2012.09.010>

Peña Espinoza, E.A. (2009). *Proyectos de desarrollo*. Recuperado de: <http://www.esap.edu.co/portal/wp-content/uploads/2017/10/3-Proyectos-de-Desarrollo.pdf>

Potes López, J. y Vargas Quimbaya, J.L. (2017). *Factores que ocasionan retrasos en obras civiles en Empresas Publicas de Neiva*. (Tesis de posgrado). Universidad de Santo Tomás. Recuperado

de: <https://repository.usta.edu.co/bitstream/handle/11634/10740/Johana%20Potes->

[2017.pdf?sequence=1&isAllowed=y](#)

Recuerdo de los Santos, P (2021). Tipos de Aprendizaje en Machine Learning: Supervisado y no Supervisado. Recuperado de: <https://empresas.blogthinkbig.com/que-algoritmo-elegir-en-ml-aprendizaje/>

Rivera Esteban, V. M. (2015). Programación, planificación y control de obras de infraestructura civil en la república de Guatemala. (Tesis de pregrado). Universidad San Carlos de Guatemala.

Romero Saldaña, M (2016). Pruebas de bondad de ajuste a una distribución normal. *Revista Enfermería del Trabajo* 2016;6:3 (105-114)

RStudio (s.f). Recuperado de: <https://www.rstudio.com/products/rstudio/>

S., A., D., H., Y., M., & U., H. (2011). Research in Modeling and Simulation for Improving Construction Engineering Operations. *Journal of Construction Engineering and Management*, 137(10), 843–852. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0000288](https://doi.org/10.1061/(ASCE)CO.1943-7862.0000288)

Salas R (2004). Redes Neuronales Artificiales. Departamento de Computación Universidad de Valparaiso.

Sánchez Albán, Pablo Israel (2020). Desarrollo de controladores con redes neuronales de aprendizaje profundo aplicando tensorflow. Carrera de Ingeniería en Electrónica, Automatización y Control. Universidad de las Fuerzas Armadas ESPE. Matriz Sangolquí. Recuperado de: <http://repositorio.espe.edu.ec/handle/21000/21705>

Sinenkov, D. (2019). Aplicación de la inteligencia artificial para la gestión de decisiones en empresas internacionales. (Tesis de maestría). Universidad de Barcelona. Recuperado de: [https://comercioexterior.ub.edu/tesina/Proyectos19-20/primer\\_proyecto/Proyecto\\_SinenkovSergei.pdf](https://comercioexterior.ub.edu/tesina/Proyectos19-20/primer_proyecto/Proyecto_SinenkovSergei.pdf).

Sthory, E. (2019). Sistema de Redes Neuronales para la Evaluación de Programas de Postgrado.(Tesis de Maestría). Universidad de Carabobo.

T. Bock. (2015). The future of construction automation: technological disruption and the upcoming ubiquity of robotics.

Tamayo, M. (2007). Tipos de Investigación.

Thomas, H. (2018). Inteligencia artificial. Revista de la Universidad Bolivariana, vol. 1 (2), pp. 1-23. Recuperado de: <https://www.redalyc.org/pdf/305/30500219.pdf>

Universidad de Jaén. (2015). Modelo de regresión lineal. Recuperado de: <http://www4.ujaen.es/~dmontoro/Metodos/Tema%209.pdf>

Universidad Nebrija. (2009). Introducción a la Inteligencia Artificial.

Vahid, S (2019). Python Machine Learning.

Wang, X., Love, P. E. D., Kim, M. J., Park, C.-S., Sing, C.-P., & Hou, L. (2013). A conceptual framework for integrating building information modeling with augmented reality. Automation in Construction, 34, 37–44. <https://doi.org/https://doi.org/10.1016/j.autcon.2012.10.012>

Webster, M. (2016). ARC Launches SKYSITE, Brings IoT to Construction Industry \_ WorldIndustrial Reporter.

## 9. Anexos

### 9.1 Anexo 1 - (Variables categóricas Ciudad y Zona) - Base de datos General

	ciudad	Total_registros
1	BOGOTÁ, D.C.	28.439
2	SOACHA	2.124
3	CHÍA	1.390
4	CAJICÁ	1.254
5	FUSAGASUGÁ	1.015
6	ZIPAQUIRÁ	763
7	TOCANCIPÁ	723
8	MOSQUERA	692
9	MADRID	691
10	COTA	437
11	GIRARDOT	395
12	RICAURTE	388
13	FUNZA	387
14	LA MESA	364

15	FACATATIVÁ	360
16	LA CALERA	221
17	ANAPOIMA	154
18	SOPÓ	150
19	VILLETA	139
20	TENJO	124
21	LA VEGA	90
22	TABIO	51
23	GACHANCIPÁ	46
24	SIBATÉ	41
25	APULO	37

	<b>Zona</b>	<b>Total registros</b>
1	10 - MULTICENTRO	3.694
2	05 - CEDRITOS	2.395
3	44 - SOACHA	2.124
4	12 - CHICO	2.119
5	02 - MAZUREN	1.667
6	36 - CORDOBA	1.627
7	54 - CHIA	1.390
8	49 - CAJICA	1.254
9	25 - AMERICAS	1.210
10	01 - MARANTA	1.109
11	65 - FUSA	1.015
12	15 - TEUSAQUILLO	983
13	16 - CHAPINERO ALTO	924
14	30 - PABLO VI	868
15	53 - ZIPAQUIRA	763
16	42 - FONTIBON	727
17	115 - TOCANCIPÁ	723
18	17 - CENTRO	723
19	59 - MOSQUERA	692
20	62 - MADRID	691
21	31 - MODELIA	677
22	34 - BOCHICA	644
23	41 - TIBABUYES	538
24	09 - LA CAROLINA	533
25	27 - CENTRAL DE ABASTOS	522
26	14 - CHAPINERO	493
27	32 - MINUTO DE DIOS	438
28	60 - COTA	437

29	13 - ROSALES	417
30	20 - QUIROGA	417
31	40 - LA CAMPIÑA	417
32	22 - MARRUECOS	405
33	229 - GIRARDOT	395
34	114- RICAURTE	388
35	37 - COLINA CAMPESTRE	388
36	58 - FUNZA	387
37	19 - 20 DE JULIO	380
38	227 - LA MESA	364
39	61 - FACATATIVA	360
40	24 - AUTOPISTA SUR	332
41	07 - LA CALLEJA	317
42	04 - CEDRO GOLF	303
43	29 - CUNDINAMARCA	301
44	33 - METROPOLIS	294
45	39 - COLINAS DE SUBA	293
46	28 - PRIMERO DE MAYO	292
47	38 - NIZA NORTE	248
48	35 - LOS ANDES	247
49	43 - BOSA	232
50	21 - VILLA DE LOS ALPES	227
51	56 - LA CALERA	221
52	26 - KENNEDY	198
53	23 - TUNJUELITO	171
54	06 - ATABANZA	158
55	116 - ANAPOIMA	154
56	03 - BOSQUE DE PINOS	153
57	113 - SOPO	150
58	120 - VILLETA	139
59	48 - TENJO	124
60	08 - BOSQUE MEDINA	123
61	18 - SAN FACON	122
62	11 - SANTA ANA	113
63	67 LA VEGA	90
64	66 TABIO	51
65	230-GACHANCIPA	46
66	45 - SIBATE	41
67	68 APULO	37

9.2 Anexo 2: Niveles (Variables categóricas Estrato y Barrio) - Base de datos General

	<b>Barrio</b>	<b>Total registros</b>
1	20 DE JULIO	1
2	7 DE AGOSTO	3
3	ABRAHAN LINCON	4
4	ACACIAS USAQUEN	220
5	ACEVEDO TEJADA	9
6	AEROPUERTO EL DORADO	1
7	ALAMEDA	49
8	ALAMOS	45
9	ALCALA	4
10	ALCALA SPRING NORTE	5
11	ALCAZARES	8
12	ALEJANDRIA	187
13	ALFONSO LOPEZ	57
14	ALGARRA	31
15	ALGARRA 3	23
16	ALGARRA III	11
17	ALGARRA LAS VILLAS	11
18	ALPES DE LA MACARENA	4
19	ALSACIA	26
20	ALTA BLANCA	14
21	ALTICO	27
22	ALTO DE LOS LAGARTOS	8
23	ALTOS DE CHOZICA	71
24	ALTOS DE LA VIRGEN	3
25	ALTOS SOTILEZA	8
26	ANDALUCIA II	25
27	ANDES	8
28	ANDES NORTE	57
29	ANGELES TEJAR	2
30	ANTIGUA VÍA CHÍA	6
31	ANTIGUO COUNTRY	205
32	ANTONIO NARIÑO	2
33	ARGELIA	2
34	ARRAYANES	6
35	ASUNCION	8
36	AUTOPISTA MUZU ORIENTAL	2
37	AUTOPISTA SUR	2
38	AV CHILACOS	40
39	BAHIA SOLANO	18

40	BALMORAL	19
41	BANCO CENTRAL	13
42	BARANDILLA	3
43	BARRANCAS	77
44	BATAN	137
45	BATAVIA	4
46	BAVARIA 1 SECTOR	2
47	BELALCAZAR	49
48	BELEN	25
49	BELEN FONTIBON	9
50	BELLA SUIZA	261
51	BELLAVISTA	32
52	BELLO HORIZONTE	4
53	BELMIRA	49
54	BENJAMIN HERRERA	9
55	BERLÍN	22
56	BETANIA	3
57	BILBAO	55
58	BOHIOS	18
59	BOLIVIA	54
60	BOLONIA	3
61	BOSA BRASIL	21
62	BOSA EL PORVENIR	2
63	BOSA EL RECREO	45
64	BOSA INDEPENDENCIA	2
65	BOSA LA INDEPENDENCIA	15
66	BOSA LA LIBERTAD	5
67	BOSA NOVA	18
68	BOSA NOVA EL PORVENIR	5
69	BOSA PALESTINA	18
70	BOSA RECREO	32
71	BOSA SAN JOSE	18
72	BOSQUE ALTO	4
73	BOSQUE CALDERON	72
74	BOSQUE DE PINOS I	40
75	BOSQUE IZQUIERDO	4
76	BOSQUE POPULAR	1
77	BOSQUES DE SAN JORGE	11
78	BOSTON	6
79	BOYACA REAL	47
80	BRAVO PAEZ	11
81	BRAZUELOS	25



82	BRISAS ALDEA-FONTIBON	2
83	BRITALIA	97
84	BRITALIA NORTE	41
85	BUENA SUERTE	3
86	BUENA VISTA	1
87	BUENOS AIRES	26
88	CABAÑA	16
89	CAGUA	37
90	CALATRABA	19
91	CALLEJA BAJA	10
92	CALLEJON SANTA BARBARA	2
93	CALVO SUR	6
94	CAMINO AL META	2
95	CAMPANELLA	18
96	CAMPIN OCCIDENTAL	34
97	CAMPINCITO	7
98	CAMPIÑA DE SUBA	8
99	CANADA GUIRA	9
100	CANDELARIA LA NUEVA	52
101	CANELON	9
102	CANODROMO	8
103	CANTAGALLO	24
104	CANTALEJO	19
105	CAOBOS SALAZAR	50
106	CAPELLANIA	33
107	CAPELLANÍA	17
108	CAPRI	13
109	CASABLANCA SUBA	32
110	CASTILLA	226
111	CEDRITOS	848
112	CEDRITOS CAPRI	2
113	CEDRO NARVAEZ	43
114	CEDRO SALAZAR	102
115	CENTRAL DE MEZCLAS	2
116	CENTRO	53
117	CENTRO ADMINISTRATIVO	3
118	CENTRO ENGATIVA	6
119	CENTRO ENGATIVA II	2
120	CENTRO HISTORICO	12
121	CENTRO INDUSTRIAL	5
122	CENTRO USME	1
123	CERRO DE SUBA	9

124	CHANIZADA	2
125	CHAPINERO	150
126	CHAPINERO ALTO	74
127	CHAPINERO CENTRAL	133
128	CHAPINERO NORTE	18
129	CHAPINERO OCCIDENTAL	3
130	CHICALA	8
131	chico	93
132	CHICO NAVARRA	71
133	CHICO NAVARRO	57
134	CHICO NORTE	352
135	CHICO NORTE II SECTOR	134
136	CHICO NORTE III SECTOR	131
137	CHICO RESERVADO	22
138	CIUDAD BERNA	26
139	CIUDAD BOLIVAR LUCERO BAJO	20
140	CIUDAD DEL SOL	25
141	CIUDAD JARDÍN	4
142	CIUDAD JARDIN NORTE	34
143	CIUDAD JARDIN SUR	52
144	CIUDAD SALITRE	16
145	CIUDAD SALITRE NOR-ORIENTAL	29
146	CIUDAD SALITRE SUR-ORIENTAL	48
147	CIUDAD VERDE	666
148	CIUADAELA COLSUBSIDIO	8
149	CIUADAELA COLSUBSIDIO MAIPORE	22
150	CIUADAELA COLSUBSIDIO MAIPORÉ	3
151	CIUADAELA LA PROSPERIDAD	41
152	CIUADAELA LA TOSCANA	14
153	CLARET	12
154	CLASS ROMA	33
155	CLUB DE LOS LAGARTOS	17
156	COBURGO	14
157	COLINA CAMPESTRE	73
158	COLINA DE SAN RAFAEL	25
159	COLINAS DE SUBA	70
160	COMPARTIR	23
161	COMUNA 3	8
162	CONCEPCION NORTE	27
163	CONEJERA	8
164	CONJUNTO AQUA CONDOMINIO	2
165	CONTADOR	61

166	COOTRADECUN	2
167	COSTA RICA	3
168	COUNTRY	4
169	COUNTRY CLUB	8
170	CUNDINAMARCA	10
171	DELICIAS	7
172	DELICIAS NORTE	2
173	DIAMANTE	5
174	DIAMANTE ORIENTAL	10
175	DINTALITO	18
176	DIVINO SALVADOR	2
177	DUCALES	11
178	EDUARDO SANTOS	14
179	EGIPTO	14
180	EL BATAN	32
181	EL BOHIO	6
182	EL CAMPIN	24
183	EL CARIBE	11
184	EL CARMELO	6
185	EL CARMEN	13
186	EL CARMEN DE FONTIBON	26
187	EL CASTILLO	5
188	EL CHARCO	1
189	EL CHICO	352
190	EL CHIRCAL SUR	3
191	EL CONTADOR	93
192	EL CORTIJO	30
193	EL CORZO	2
194	EL ENCANTO	8
195	EL ENSUENO	20
196	EL ESCALLON	2
197	EL HATO	4
198	EL LISTON	96
199	EL LUCERO	1
200	EL MINUTO DE DIOS	80
201	EL MOCHUELO	9
202	EL NOGAL	102
203	EL PARAISO	3
204	EL PEÑÓN	10
205	EL PESEBRE	7
206	EL PINO	75
207	EL PLAN	55

208	EL PLAYON	5
209	EL POA	17
210	EL PORVENIR	4
211	EL RAMAJAL	1
212	EL REAL	9
213	EL RECODO	22
214	EL RECUERDO	14
215	EL REFUGIO	74
216	EL RETIRO	51
217	EL ROCIO CENTRO	14
218	EL ROCIO NORTE	9
219	EL ROSARIO	8
220	EL RUBI	50
221	EL SALITRE	43
222	EL SOCIEGO	34
223	EL TINTAL	109
224	EL TINTAL A.S.D.	2
225	EL TINTAL CENTRAL	9
226	EL TOBERIN	5
227	EL TRIANGULO	4
228	EL TRIUNFO	2
229	EL VERGEL	25
230	EL VERGEL ORIENTAL	11
231	EL VERVENAL	8
232	ELPEÑON	4
233	EMAUS	16
234	EMILIO SIERRA	9
235	ENCANTO	4
236	ENGATIVA ZONA URBANA	2
237	ESCOCIA	2
238	ESCUELA MILITAR	8
239	ESPARTILLAL	6
240	ESTADOS UNIDOS	2
241	ESTORIL	72
242	ESTRELLA DEL NORTE	33
243	FAENA	19
244	FATIMA	15
245	FERROCAJA FONTIBON	29
246	FLORALIA	2
247	FLORENCIA	47
248	FLORIDA BLANCA	5
249	FONTIBON CENTRO	13

250	FRANCISCO MIRANDA	6
251	FRANCO	4
252	FUNZA CENTRO	12
253	FUSACATAN	11
254	GAITAN	8
255	GALAN	4
256	GALERIAS	70
257	GALICIA	47
258	GARCES NAVAS ORIENTAL	4
259	GILMAR	128
260	GINEBRA	10
261	GRAN AMERICA	101
262	GRAN GRANADA	240
263	GRAN VÍA	2
264	GRANADA	38
265	GRANADA NORTE	15
266	GRANADA SUR	5
267	GRANCOLOMBIA	32
268	GRANJAS DEL DORADO	12
269	GRANJAS SAN PABLO	6
270	GRATAMIRA	57
271	GUADUAL FONTIBON	7
272	GUAIMARAL - CENTRO COMERCIAL BIMA	2
273	GUALOCHE	4
274	GUAYMARAL	1
275	HACIENDA CASA BLANCA	98
276	hacienda de casa blanca	16
277	HACIENDA NARANJAL	8
278	HATO 2	4
279	HAYUELOS	29
280	HELICONAS	36
281	HIPOTECHO	7
282	HIPOTECHO OCCIDENTAL	18
283	HIPOTECHO SUR	68
284	HOGARES SOACHA	137
285	HORIZONTES NORTE	38
286	HOYA TEUSACA	7
287	IBERIA	27
288	ILARCO	14
289	INGEMAR	42
290	INGLES	9
291	ISLANDIA	19

292	ISMAEL PERDOMO	3
293	JAZMIN	3
294	JJ VARGAS	6
295	JORGE URIBE BOTERO	18
296	JOSE JOAQUIN VARGAS	9
297	JUAN PABLO - NUEVA GRANADA	2
298	JUAN REY	32
299	JULIO CARO	22
300	JULIO FLOREZ	52
301	KENNEDY	5
302	KENNEDY CENTRAL	10
303	KM 5 VIA CAJICA	15
304	LA ALAMEDA	14
305	LA ALBORADA	6
306	LA ARBOLEDA	12
307	LA AURORA	27
308	LA CABANA	6
309	LA CABRERA	38
310	LA CALLEJA	133
311	LA CALLEJA ALTA	8
312	LA CAMPINA	9
313	LA CAMPIÑA	4
314	La candelaria	10
315	LA CAROLINA	148
316	LA CASTELLANA	130
317	LA CATEDRAL	24
318	LA CITA	31
319	LA CLARITA	2
320	LA COLINA	11
321	LA CONCEPCION NORTE	7
322	LA CONCORDIA	13
323	LA ESMERALDA	2
324	LA ESPAÑOLA	1
325	LA ESPERANZA	47
326	LA ESPERANZA NORTE	72
327	LA ESTACION-BOSA	11
328	LA ESTANCIA	10
329	LA FELICIDAD	67
330	LA FISCALA	3
331	LA FLORIDA	6
332	LA FLORIDA OCCIDENTAL	2
333	LA FRAGUA	2

334	LA FRAGUITA	55
335	LA GIRALDA	16
336	LA GRANJA	58
337	LA GRANJITA	48
338	LA GUACA	4
339	LA IGUALDAD	9
340	LA INDEPENDENCIA	2
341	LA LAGUNA FONTIBON	17
342	LA LIBERIA	3
343	LA LOMITA	23
344	LA MACARENA	15
345	LA MARIA	3
346	LA MERCED	25
347	LA PALMA	17
348	LA PAMPA	3
349	LA PAZ BOSA	6
350	LA PEPITA	6
351	LA PICOTA SUR	54
352	LA PRADERA NORTE	81
353	LA PRIMAVERA I	37
354	LA RESURRECCION	15
355	LA SABANA	6
356	LA SALLE	54
357	LA SOLEDAD	38
358	LA SOLEDAD NORTE	7
359	LA SONORA	1
360	LA URIBE	58
361	LA VENTA	12
362	LA VEREDA GUAVILAN	7
363	LA VEREDITA	10
364	LA VIRGEN	5
365	LAGO GAITAN	4
366	LAGOS DE CASTILLA	4
367	LAGOS DE CORDOBA	23
368	LAGOS DE TORCA	15
369	LAS ACACIAS	30
370	LAS AGUAS	14
371	LAS AMAPOLAS	2
372	LAS BRISAS	5
373	LAS CRUCES	21
374	LAS DOS AVENIDAS	1
375	LAS FERIAS	5

376	LAS FLORES	68
377	LAS MARGARITAS	10
378	LAS MERCEDES SUBA	14
379	LAS NIEVES	108
380	LAS QUINTAS	14
381	LAS TERRAZAS	1
382	LAS VEGAS CENTRO	2
383	LAS VENTAS	3
384	LAS VILLAS	60
385	LIBERTADOR	46
386	LISBOA	190
387	LLANO LARGO	10
388	LOMBARDIA	17
389	LOS ALAMOS	23
390	LOS ALPES	20
391	LOS ANDES	17
392	LOS CEDROS	96
393	LOS CEDROS ORIENTAL	37
394	LOS CERROS	5
395	LOS LAURELES	2
396	LOS MONJES	4
397	LOS OCOBOS	9
398	LOS PINOS	1
399	LOS ROSALES	44
400	LOS SAUCES	4
401	LOS SAUCES BAJO	11
402	MADELENA	60
403	MAIPORE	16
404	MAIZ AMARILLO	18
405	MANAS	10
406	MANILA	64
407	MARANTA	20
408	MARCELLA	9
409	MARCO FIDEL SUAREZ	10
410	MARIA CRISTINA	73
411	MARICHUELA	3
412	MARLY	50
413	MARSELLA	53
414	MAZUREN	107
415	METROPOLIS	12
416	MIRA FLORES	11
417	MIRANDELA	25



418	MODELIA	46
419	MODELO NORTE	49
420	MOLINOS	20
421	MOLINOS NORTE	82
422	MONACO	16
423	MOSQUERA	7
424	MUEQUETA	2
425	MURILLO TORO	2
426	MUZÚ	1
427	NARINO SUR	3
428	NAVARRA	3
429	NICOLAS DE FEDERMAN	181
430	NIZA	27
431	NIZA NORTE	35
432	NIZA SUBA	51
433	NIZA SUR	6
434	NO ASIGNADO	12357
435	NORMANDIA	51
436	NORMANDIA OCCIDENTAL	8
437	NUESTRA SEÑORA DEL CARMEN	1
438	NUEVA AUTO	27
439	NUEVA AUTOPISTA	56
440	NUEVA CASTILLA	92
441	NUEVA DELI	34
442	NUEVA ZELANDIA	6
443	NUEVO COUNTRY	30
444	NUEVO TECHO	29
445	OLAYA	73
446	OLIVOS	8
447	ORQUIDEAS	8
448	ORTEZAL	20
449	OSORIO	159
450	OTRO	4820
451	PALERMO	26
452	PALERMO SUR	3
453	PALOQUEMAO	19
454	PARAMO	19
455	PARCELA EL PORVENIR	2
456	PARCELACION TERMALES	5
457	PARDO RUBIO	19
458	PARIS	5
459	PARQUE LAGARTOS	3

460	PASADENA	247
461	PASO ANCHO	4
462	PATIO BONITO	6
463	PATRIA	1
464	PEDREGAL	4
465	PEKIN	41
466	PEÑALISA	11
467	PEPINITOS	1
468	PERDOMO	9
469	PINAR DE SUBA	25
470	PINARES	10
471	PINOS DE LOMBARDIA	33
472	PIO XII	10
473	PLANADAS	57
474	POLO CLUB	14
475	PONTEVEDRA	48
476	POPULAR MODELO	19
477	PORCIUNCULA	20
478	PORTACHUELO	18
479	PORTALES DEL NORTE	140
480	PORVENIR	18
481	POTOSI	25
482	POZO AZUL	2
483	PRADO PINZON	70
484	PRADO VERANIEGO	83
485	PRADO VERANIEGO NORTE	12
486	PRADO VERANIEGO SUR	4
487	PRADOS DEL MIRADOR	9
488	PRIMAVERA II	3
489	PRIMAVERA OCCIDENTAL	4
490	PROVIVIENDA ORIENTAL	3
491	PUEBLO VIEJO	40
492	PUENTE ARANDA	17
493	PUENTE GRANDE	27
494	PUENTE LARGO	165
495	PUERTA DE TEJA	11
496	PUERTA DE TEJO	23
497	PUERTO RICO	6
498	QUESADA	20
499	QUINTA CAMACHO	57
500	QUINTA PAREDES	77
501	QUINTANARES	14

502	QUIROGA	15
503	RAMIREZ	5
504	REDIL	4
505	REFUGIO ALTO	6
506	RENANIA URAPANES	2
507	RESERVA DE MADRID	6
508	RESTREPO OCCIDENTAL	27
509	RICAURTE	35
510	RINCON DE LA VALVANERA	3
511	RINCON DE SANTA INES	6
512	RINCON DE SANTAFE	10
513	RINCON DE SUBA	4
514	RINCON DEL CHICO	100
515	RINCÓN DEL DIAMANTE	4
516	RINCON DEL PUENTE	21
517	RINCON SANTO	6
518	RIO FRIO	7
519	RÍO FRÍO	67
520	ROCIO	7
521	ROSA BLANCA	22
522	ROSALES	28
523	ROSALES ALTO	16
524	ROSARIO	5
525	SABANA DE TIBABUYES NORTE	10
526	SAGRADO CORAZON	18
527	SALITRE	44
528	SALITRE OCCIDENTAL	40
529	SALITRE SUBA	9
530	SAMANA	2
531	SAMPER	30
532	SAMPER MENDOZA	26
533	SAN ANDRESITO 38	17
534	SAN ANTONIO	26
535	SAN ANTONIO NOROCCIDENTAL	12
536	SAN ANTONIO NORTE	33
537	SAN BERNARDINO	22
538	SAN BERNARDINO XVI	10
539	SAN BERNARDINO XVIII	1
540	SAN BLAS II	21
541	SAN CRISTOBAL NORTE	49
542	SAN CRISTOBAL SUR	27
543	SAN DIEGO	63

544	SAN DIEGO-BOSA	4
545	SAN EUSEBIO	5
546	SAN FERNANDO	30
547	SAN FRANCISCO PARTE ALTA	4
548	SAN GABRIEL NORTE	46
549	SAN HUMBERTO	7
550	SAN IGNACIO	23
551	SAN ISIDRO	9
552	SAN JOAQUIN	2
553	SAN JOSE DE BAVARIA	30
554	SAN JOSE DE FONTIBON	10
555	SAN JOSE DE LA ESTRADA	2
556	SAN JOSE DE USAQUEN	41
557	SAN JOSE DEL PRADO	50
558	SAN JOSE SPRING	3
559	SAN JOSE V SECTOR	3
560	SAN JUANITO	2
561	SAN LUIS	132
562	SAN MARTIN	99
563	SAN MATEO	40
564	SAN MIGUEL	13
565	SAN PABLO	4
566	SAN PABLO BOSA	2
567	SAN PABLO JERICO	9
568	SAN PATRICIO	417
569	SAN PEDRO	6
570	SAN RAFAEL	15
571	SAN VICENTE SUR ORIENTAL	3
572	SAN VICTORINO	5
573	SANTA ANA	70
574	SANTA ANA OCCIDENTAL	55
575	SANTA ANA ORIENTAL	5
576	SANTA ANA SUR	1
577	SANTA ANITA	16
578	SANTA BARBARA	518
579	SANTA BARBARA BAJA	11
580	SANTA BARBARA CENTRAL	122
581	SANTA BARBARA OCCIDENTAL	282
582	SANTA BARBARA ORIENTAL	225
583	SANTA BIBIANA	248
584	SANTA CATALINA	6
585	SANTA CECILIA	2

586	SANTA FE BOSA	3
587	SANTA HELENA	23
588	SANTA HELENITA	1
589	SANTA INES	7
590	SANTA ISABEL	17
591	SANTA LUCIA	3
592	SANTA MARGARITA	85
593	SANTA MARIA	18
594	SANTA MARIA DEL LAGO	25
595	SANTA MATILDE	14
596	SANTA PAULA	43
597	SANTA PAULA ALTA	77
598	SANTA RITA	5
599	SANTA RITA SUR ORIENTAL	14
600	SANTA ROSA	119
601	SANTA ROSITA	10
602	SANTA TERESA	132
603	SANTA TERESITA	7
604	SANTANDER	29
605	SANTANDER - LA GIRALDA	3
606	SANTO DOMINGO	6
607	SARA LUCÍA	1
608	SAUZAL	7
609	SAUZALITO	5
610	SEC STA INES	1
611	SECTOR BOHIO	13
612	SECTOR LA PALMA	12
613	SEMINARIO	2
614	SIETE TROJES	10
615	SIMON BOLIVAR	5
616	SIXTO LLERAS LOPEZ	34
617	SOCIEGO	4
618	SOTAVENTO	9
619	SOTILEZA	21
620	SPRING	22
621	SUBA - RINCON	4
622	SUBA / FONTANAR DEL RIO	2
623	SUBA CAMPIÑA	3
624	SUBA CENTRO	20
625	SUBA FONTANAR	24
626	SUBA LOMBARDIA	2
627	SUBA PROVENZA	23

628	SUBA RINCON	5
629	SUBA SALITRE	16
630	SUBA URBANO	49
631	SUCRE	66
632	SURAMERICA	27
633	SUTILEZA RESERVADO	20
634	TEUSAQUILLO	175
635	TIBABITA	225
636	TIBABUYES	42
637	TIBANA	28
638	TIERRA BUENA	48
639	TIERRA LINDA	9
640	TIERRA LINDA SAN RAFAEL	14
641	TIGUA SAN JOSÉ	4
642	TIMIZA	11
643	TIMIZA A	15
644	TOSCANA	1
645	TRANQUILANDIA	14
646	TRES ESQUINAS	5
647	TUNA	2
648	TUNA ALTA	6
649	TUNA BAJA	24
650	TUNAL ORIENTAL	18
651	TUNJUELITO	3
652	Unicentro	7
653	USAQUEN	53
654	USME	14
655	USME RURAL	1
656	VALLADOLID	13
657	VALLES DE CAFAM	21
658	VALVANERA	1
659	VARIANTE CHÍA - CAJICÁ	16
660	VARIANTE CHÍA - COTA	36
661	VDA ALTAMAR	3
662	VDA BOJACA	18
663	VDA CALAHORRA	15
664	VDA CANELON	4
665	VDA CHUNTAME	18
666	VDA CHUTAME	3
667	VDA EL DIAMANTE	5
668	VDA LA Balsa	19
669	VDA LA ESMERALDA	11

670	VDA LA FUENTE	1
671	VDA LA PALMA	6
672	VDA LA PUNTA	1
673	VDA SAN JOSE	13
674	VDA SAN JUDAS	2
675	VDA SIETE TROJES	2
676	VDA SINDAMANOY	15
677	VDA VERGANZO	3
678	VDA ZARAGOZA	12
679	VEGAS DE SANTA ANA	28
680	VELODROMO	19
681	VENECIA	2
682	VERACRUZ	30
683	VERAGUAS	10
684	VERAGUAS CENTRAL	13
685	VEREDA AGUABLANCA	26
686	VEREDA BOJACA	57
687	VEREDA BULUCAIMA	2
688	VEREDA CALAHORRA	21
689	VEREDA CANAVITA	16
690	VEREDA CANELO	3
691	VEREDA CANELON	30
692	VEREDA CANELÓN	26
693	VEREDA CENTRO ENGATIVA	1
694	VEREDA CERCA DE PIEDRA	2
695	VEREDA CHUNTAME	31
696	VEREDA CHUZACA	3
697	VEREDA COGUA	5
698	VEREDA CUCHARAL	1
699	VEREDA DE SAN JOSE	11
700	VEREDA EL ABRA	31
701	VEREDA EL HATO	30
702	VEREDA EL TINTAL II	4
703	VEREDA FONQUETÁ	13
704	VEREDA GUACAMAYAS	2
705	VEREDA LA BALSA	13
706	VEREDA LA CHICA	13
707	VEREDA LA TRINIDAD	16
708	VEREDA LAGUNA VERDE	1
709	VEREDA LAS MERCEDES	21
710	VEREDA MANUELES CASA GRANDE	3
711	VEREDA MEUSA	16

712	VEREDA PUEBLO VIEJO	35
713	VEREDA SAN JORGE	13
714	VEREDA SAN JUAN	8
715	VEREDA SANTA BÁRBARA	1
716	VEREDA SANTA ISABEL	1
717	VEREDA SANTA ISABEL DE POTOSI	13
718	VEREDA SANTA SOFIA	4
719	VEREDA SIBERIA	2
720	VEREDA SUBA CERROS	16
721	VEREDA SUBA CERROS II	21
722	VEREDA TIBABUYES	19
723	VEREDA TIBITO	11
724	VEREDA TRINIDAD	10
725	VEREDA VERGANZO	220
726	VEREDA VIOLETA	3
727	VEREDA VUELTA GRANDE	9
728	VEREDA VUELTA LA TRAMPA	2
729	VERGANZO	3
730	VERGEL OCCIDENTAL	20
731	VIA AGUA BLANCA	15
732	VÍA AGUA BLANCA	5
733	VIA ARGENTINA	3
734	VÍA CAJICÁ - TABIO	8
735	VÍA CAJICÁ - ZIPAQUIRÁ	55
736	VIA CANELON TABIO	6
737	VIA INDUMIL	20
738	VIA NARIÑO	13
739	VÍA SAN JOAQUÍN	2
740	VÍA SASAIMA	6
741	VICTORIA NORTE	54
742	VILLA ALEMANA	2
743	VILLA ALSACIA	162
744	VILLA ANNY	14
745	VILLA CECILIA	2
746	VILLA CLAUDIA	9
747	VILLA DEL MAR	4
748	VILLA DEL PRADO	14
749	VILLA DEL RIO	24
750	VILLA ELISA	3
751	VILLA GLADYS	24
752	VILLA HERMOSA	7
753	VILLA JAVIER	8



754	VILLA MARIA	1
755	VILLA MAYOR ORIENTAL	32
756	VILLA NELLY	2
757	VILLAMAR	12
758	VILLAS DE GRANADA	56
759	VILLAS DE GRANADA I	26
760	VILLAS DEL PRADO	2
761	VILLEMAR	75
762	YOMASA	12
763	ZUAME	2

