



UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA Y
CIENCIAS APLICADAS

MODELO PARA EXPLICAR VALOR
DEL SUELO PARA PROYECTOS
INMOBILIARIOS EN LA COMUNA DE
LAS CONDES

MATÍAS SANTIAGO GUALDA MOREL

MEMORIA PARA OPTAR AL TITULO DE
INGENIERO CIVIL INDUSTRIAL

PROFESOR GUÍA: SANTIAGO TRUFFA

SANTIAGO, NOVIEMBRE DE 28

UNIVERSIDAD DE LOS ANDES
FACULTAD DE INGENIERÍA Y
CIENCIAS APLICADAS

AGRADECIMIENTOS

En primer lugar, quiero agradecer a mis padres y a mis hermanos. Quienes han estado a mi lado en toda mi etapa universitaria, por darme su constante apoyo en las decisiones que tomé.

En segundo lugar, quiero agradecer a mi profesor guía, Christian Wegmann, por darme la oportunidad de trabajar en la memoria con él, por su disposición y constancia en juntarnos a trabajar cada semana.

En tercer lugar, quiero agradecer a Santiago Truffa, por darme su apoyo en este trabajo y otorgarme las herramientas necesarias.

En cuarto lugar, quiero agradecer a Orlando Contreras, del centro de escritura, quien tuvo la paciencia y entrega para revisar la memoria.

En quinto lugar, quiero agradecer a mis amigos y compañeros de la universidad, quienes me acompañaron en esta gran etapa, también por sus ayudas y consejo para mi desarrollo en la universidad.

Por último, quiero agradecer a la Facultad y sus profesores, por su disposición a enseñar y otorgarnos los conocimientos y permitirme completar esta etapa.

ÍNDICE GENERAL

RESUMEN	VIII
1. INTRODUCCIÓN.....	1
1.1 MOTIVACIÓN.....	1
1.2 PROBLEMA.....	3
1.3 OBJETIVOS DE LA MEMORIA.....	4
1.3.1 OBJETIVO GENERAL.....	4
1.3.2 OBJETIVOS ESPECÍFICOS	4
1.4 METODOLOGÍA	5
1.4.1 Tareas asociadas con primer objetivo.....	5
1.4.2 Tareas asociadas con segundo objetivo	5
1.4.3 Tareas asociadas con tercer objetivo	6
1.4.4 Tareas asociadas con cuarto objetivo	6
1.4.5 Tareas asociadas con quinto objetivo	6
1.5 ALCANCES	6
2. MARCO TEÓRICO Y REVISIÓN BIBLIOGRÁFICA	8
2.1 IMPORTANCIA DEL PRECIO DE SUELO	8
2.2 LA UTILIDAD DEL CÁLCULO DEL PRECIO DEL SUELO PARA LAS EMPRESAS INMOBILIARIAS EN SUS PROYECTOS	9
2.3 LA DINÁMICA ENTRE PRECIOS DE SUELO Y VIVIENDA	10
2.4 MODELO UTILIZADO PARA PRECIOS DE VIVIENDA.....	11
2.5 EL IMPACTO DE LA PLANIFICACIÓN URBANA EN EL PRECIO DEL SUELO	11
2.6 REGRESION LINEAL	12
2.7 REGRESIÓN LINEAL EN RSTUDIO.....	13
2.8 CARET Y METODO LVQ	14
2.9 HETEROCEDASTICIDAD	16
2.10 MULTICOLINEALIDAD.....	17
2.11 AUTOCORRELACION.....	18
3. LEVANTAMIENTO DE DATOS Y GENERACIÓN DE BASE DE DATOS ..	19

3.1	LEVANTAMIENTO DE DATOS	19
3.1.1	Levantamiento de permisos e identificación de proyectos	19
3.1.2	Identificación de roles.....	21
3.1.3	Levantamiento de transacciones.....	22
3.2	ÁNÁLISIS DE DATOS	25
3.2.1	Creación base de datos.....	25
3.2.2	Variables independientes	26
4.	TRANSFORMACIÓN Y ANALISIS DE RESULTADOS.....	29
4.1	TRANSFORMACIÓN VARIABLES	29
4.2	CORRELACIÓN DATOS	34
4.3	AJUSTE REGRESIÓN LINEAL MÚLTIPLE	36
4.4	HETEROCEDASTICIDAD EN LOS DATOS.....	44
4.5	ANÁLISIS DE RESULTADOS REGRESION LINEAL.....	45
4.6	ANÁLISIS DE AUTOCORRELACIÓN Y MULTICOLINEALIDAD.....	51
5.	CONCLUSIONES Y RECOMENDACIONES.....	54
5.1	CONCLUSIONES GENERALES	54
5.2	CONCLUSIONES ESPECÍFICAS.....	55
5.3	RECOMENDACIONES FUTURAS.....	56
6.	BIBLIOGRAFÍA.....	59
7.	ANEXOS.....	61
	ANEXO 1. PERMISOS DE EDIFICACIÓN LEVANTADOS.....	61
	ANEXO 2. DOCUMENTOS DE TRANSACCIÓN.....	61
	ANEXO 3. RESULTADO PAQUETE DE CARET	62
	ANEXO 4. PRIMERA REGRESIÓN LINEAL MÚLTIPLE.....	62
	ANEXO 5. REGRESIÓN LINEAL DESPUÉS DE ELIMINAR VARIABLES INDEPENDIENTES.....	63
	ANEXO 6. REGRESIÓN LINEAL CON VARIABLES TRANSFORMADAS	63
	ANEXO 7. RESULTADO CARET CON VARIABLES ELEGIDAS	65
	Anexo 8. PRUEBA DURBIN-WATSON.....	65

ANEXO 9. RESULTADO RESIDUOS REGRESIÓN LINEAL	66
ANEXO 10. DESVIACION ESTANDAR COEFICIENTE DE CONSTRUCTIBILIDAD	66
ANEXO 11. DESVIACION ESTANDAR DENSIDAD	66
ANEXO 12. COMPARACIÓN VALORES REALES VS VALORES PREDICTIVOS DE LA REGRESIÓN LINEAL	67

ÍNDICE DE TABLAS

TABLA 1: PERMISOS DE EDIFICACIÓN LEVANTADOS.	17
TABLA 2: ROLES IDENTIFICADOS.	18
TABLA 3: DOCUMENTOS DE TRANSACCIÓN.	20
TABLA 4: ESTADÍSTICAS VALOR UNITARIO EN UF.	21
TABLA 5: EJEMPLO DE TRANSFORMACIONES.....	27
TABLA 6: RESUMEN VARIABLE Y SUS TRANSFORMACIONES.....	27
TABLA 7: PRIMERA REGRESION LINEAL MÚLTIPLE.	34
TABLA 8: PRIMERA REGRESION LINEAL MÚLTIPLE RESULTADOS.	35
TABLA 9: REGRESIÓN LINEAL DESPUES DE ELIMINAR VARIABLES.	38
TABLA 10: REGRESION LINEA DESPUES DE ELIMINAR VARIABLES RESULTADOS.	39
TABLA 11: REGRESION LINEAL CON VARIABLES TRANSFORMADAS.....	39
TABLA 12: REGRESION LINEAL CON VARIABLES TRANSFORMADAS RESULTADOS.	40
TABLA 13: TEST BREUSCH-PAGAN.	41
TABLA 14: REGRESION LINEAL METODO ESTIMACION ROBUSTOS.	41
TABLA 15: REGRESION LINEAL METODO ESTIMACION ROBUSTOS RESULTADOS.	42
TABLA 16: PRUEBA DURBIN-WATSON.	50

ÍNDICE DE ILUSTRACIONES

ILUSTRACIÓN 1 GRÁFICO VALOR M2 EN UF.....	22
ILUSTRACIÓN 2 HISTOGRAMA VALOR METRO CUADRADO.....	25
ILUSTRACIÓN 3 HISTOGRAMA VARIABLE DEPENDIENTE EN LOGARITMO. ..	25
ILUSTRACIÓN 4 MATRIZ DE CORRELACIÓN SUPERFICIE EDIFICADA	28

RESUMEN

Las inmobiliarias enfrentan desafíos al carecer de un modelo predictivo sólido para, lo que puede afectar su rentabilidad y éxito a largo plazo. La ubicación, las restricciones de edificación y el estado del suelo son factores clave que influyen en los precios. La falta de información precisa sobre precios de viviendas en Chile crea incertidumbre en el mercado. El crecimiento del mercado inmobiliario chileno trae consigo desafíos adicionales, como el aumento de costos y la disponibilidad de mano de obra.

Es por estos problemas que en este trabajo se construyó este modelo, para predecir el precio del suelo a través de una regresión lineal múltiple. Para el levantamiento de datos se levantaron todos los permisos de edificación de la comuna de Las Condes en la página de esta. Una vez levantados todos estos permisos, se filtraron bajo ciertos criterios y se levantaron las transacciones por los terrenos. Así se dividieron los terrenos pudiendo levantar las medidas de este.

Una vez armada la base de datos se comenzó a trabajar con ella, se realizaron transformaciones a las variables para ajustar la normalidad de los datos y se eliminaron los datos que se consideraron irrelevantes con sus análisis respectivos. Una vez construida la regresión lineal, se aplicaron los modelos respectivos para verificar la veracidad de los resultados.

Se realizó un análisis de cada coeficiente de todas las variables independientes y la explicación de por qué su valor contextualizándolos a como de útil es la información a las inmobiliarias.

Se concluye que existe una influencia significativa de varias variables en el precio del suelo, incluyendo zonas censales, densidad de construcción, coeficiente de ocupación de suelo, número de pisos y la presencia de locales comerciales. Estas variables son cruciales para las inmobiliarias, ya que les permiten tomar decisiones financieras sólidas y estratégicas que afectan la rentabilidad de sus proyectos y carteras de inversión. Aunque el modelo desarrollado tiene un R cuadrado ajustado bajo, ofrece información valiosa y sólida sobre cómo estas variables influyen en el precio del suelo.

1. INTRODUCCIÓN

En este capítulo, se presentan los puntos que permiten comprender y fundamentar el proceso realizado en esta memoria. Los puntos son los siguientes: motivación, problema, alcances, objetivos y metodología.

1.1 MOTIVACIÓN

En el mundo de los bienes raíces, donde las decisiones financieras pueden tener un impacto significativo en el éxito de un proyecto inmobiliario, contar con herramientas precisas y fundamentadas es esencial.

En este informe, exploraremos en detalle cómo la regresión lineal aplicada al precio del suelo, considerando todas sus variables independientes, se ha convertido en un activo invaluable para las inmobiliarias en su estudio de mercado. Este enfoque analítico permite a las empresas del sector no solo comprender las tendencias pasadas, sino también anticipar las dinámicas futuras del mercado de terrenos, optimizando así la toma de decisiones estratégicas.

En este contexto, las inmobiliarias enfrentan desafíos que impactan su rentabilidad y éxito a largo plazo. La falta de un modelo predictivo para el precio del suelo impide la previsión de tendencias, exponiéndolas a cambios inesperados en el mercado que pueden perjudicar sus proyectos y carteras de inversión.

La carencia de información precisa sobre el precio del suelo conlleva el riesgo de asignar incorrectamente recursos financieros, lo que aumenta la posibilidad de invertir en proyectos poco rentables debido a una comprensión deficiente del mercado.

Hoy en día, persiste una incertidumbre en torno a los factores precisos que inciden en los precios del suelo en Chile. Abordar estos elementos que ejercen influencia en las tasas de suelo chilenas demanda una investigación minuciosa, con el fin de dilucidar la extensión del impacto de cada uno de estos factores en los precios del mercado.

Una variable importante en este valor es su localización, pues el entorno de un lugar afecta la calidad de vida de las personas, ya sea para vivienda o no. El entorno puede dar facilidades tanto económicas como sociales, por lo que, según las comodidades del sector, el precio puede variar. Por ejemplo, el precio de suelo para viviendas puede ser más elevado en un sector donde exista un mayor acceso a servicios públicos, económicos, colegios, mayor seguridad, etc. En cambio, en un lugar donde haya escasez de servicios tanto públicos como privados, el precio del suelo será menor.

Otro factor clave es el uso y las restricciones de edificación que definen los instrumentos de planificación territorial. Un suelo donde se limite la construcción de edificaciones de mayor altura tendrá probablemente un menor valor. También afecta el estado y superficie del suelo. Las características del estado del suelo son fundamentales a la hora de definir un precio. Un suelo con menores dificultades para edificar tendrá asociados precios más bajos, por lo tanto, el inversionista tendrá un nivel de riesgo menor al adquirir este tipo de suelo. Asimismo, si las dificultades son menores, habrá una mayor demanda por ese suelo, por lo que el precio sería mayor por este lado.

También existe la orientación de este suelo, es decir, la exposición de la luz solar y vistas del terreno, que influirán en su valor. Quienes adquieren una vivienda analizan todos los factores que complementan los terrenos, por lo que factores como el tiempo durante el cual el terreno recibe diariamente luz solar puede influir en el precio.

Los precios de los terrenos, durante los últimos 10 años, han experimentado alzas de más de 200%, según el sector. En la comuna de Las Condes, (comuna por estudiar en este trabajo) se produjo un aumento del 295% en comparación con el año 2010 al 2019, esto se debe a distintos factores que a medida que pasan los años influyen en el precio del suelo. (M, 2021)

En Chile, las viviendas representan la principal riqueza de los hogares y son esenciales para el sistema financiero como garantía de créditos hipotecarios. Los cambios en los precios de las viviendas impactan significativamente en el consumo e inversiones de los hogares, así como en la posición financiera de las entidades financieras.

Existe una falta de información en Chile sobre los precios de las viviendas y una escasez de indicadores confiables y continuos en el mercado inmobiliario. Se mencionan dos indicadores utilizados para rastrear los precios de las viviendas: transacciones registradas por el Conservador de Bienes Raíces y ofertas de venta recopiladas por la Asociación de Corredores de Propiedades (Parrado, 2005).

La elaboración de un modelo que explique cómo la normativa urbana y la ubicación afectan los precios de las viviendas presenta desafíos debido a la alta heterogeneidad de las viviendas, la falta de datos sobre transacciones anteriores y la infrecuente frecuencia de transacciones de viviendas en el mercado chileno (aproximadamente cada 9 años). Estos factores dificultan la comparación y seguimiento precisos de los precios de las viviendas en el tiempo (Parrado, 2005).

El mercado inmobiliario en Chile, especialmente en Santiago, ha experimentado un crecimiento espectacular en la última década, con un costo promedio de propiedades en la ciudad de Santiago de 225.000 dólares y un promedio nacional de 155.000 dólares. A pesar de este crecimiento, los precios de las propiedades han experimentado un aumento significativo en los últimos años debido a múltiples factores, incluyendo el aumento en los costos de materiales de construcción, la inflación resultante de la pandemia, los retiros de fondos de pensiones (AFP), la incertidumbre política y la disponibilidad reducida de mano de obra (Engel & Völkers Blog, 2015).

1.2 PROBLEMA

El valor del suelo es un factor relevante en la compra de viviendas, ya que influye directamente en sus precios. Sin embargo, en la actualidad no existe un modelo que explique de manera completa y precisa la dinámica y evolución del precio del suelo a lo largo del tiempo.

Las inmobiliarias, al enfrentarse a un mercado de terrenos en constante evolución, se encuentran en una encrucijada. La falta de un modelo sólido que explique el precio del suelo y su dinámica a lo largo del tiempo representa un problema significativo para la toma de decisiones estratégicas.

El desafío radica en las particularidades individuales de cada terreno, las cuales afectan su precio de manera única. Cada terreno es diferente y difícilmente comparable con otros. Además, existen diferentes entidades involucradas en las transacciones y valorización del suelo, lo que genera disparidades en su apreciación.

Considerando lo anterior, surge la oportunidad de desarrollar un modelo predictivo del precio del suelo, que analice los factores que lo afectan, su relevancia y cómo influyen en su valor.

1.3 OBJETIVOS DE LA MEMORIA

A continuación, se presenta el objetivo general y los específicos de la memoria, que serán desarrollados a través del escrito.

1.3.1 OBJETIVO GENERAL

El objetivo general de esta memoria es construir un modelo que analice y explique el precio del suelo utilizando herramientas econométricas, con el fin de comprender y predecir este precio, de manera que facilite a las inmobiliarias para la toma de decisiones estratégicas, contribuyendo así a mejorar la rentabilidad y el éxito a largo plazo en sus proyectos y carteras de inversión.

1.3.2 OBJETIVO ESPECÍFICOS

Para alcanzar el objetivo general, se desprende un conjunto de objetivos específicos, los que van a determinar cómo se desarrollará esta memoria. Estos objetivos son:

- a) Levantar los permisos de edificación documentados en la comuna de Las Condes entre los años 2012 y 2021 a través de la sección de transparencia activa de web municipal.
- b) Identificar permisos que corresponden a proyectos inmobiliarios y roles de avalúo asociados con dichos permisos. Identificar precios de transacción de terrenos asociados con los permisos de los proyectos inmobiliarios.

- c) Realizar una regresión con los datos anteriormente levantados para construir el modelo explicativo del precio de suelo y verificar la veracidad de los datos a través de distintos modelos.
- d) Analizar los resultados con un enfoque hacia las inmobiliarias y como de útil es esta información para la inversión en terrenos.
- e) Validar el modelo, concluir sobre el análisis y hacer propuestas para mejorar este.

1.4 METODOLOGÍA

La metodología de esta memoria está dividida en dos etapas. En la primera de ellas, se realiza un proceso de investigación en que se reúne la información necesaria y se construye la base de datos. En la segunda, se interpretan los datos obtenidos a través de la investigación realizada.

1.4.1 TAREAS ASOCIADAS CON PRIMER OBJETIVO

- a) Levantar todos los permisos de edificación tratados en los años 2012 al 2021 en la Página de transparencia de la comuna de Las Condes.
- b) Construir la base de datos de los permisos, dividiéndolos por sus códigos de comuna, manzana y predio, juntos con sus fechas respectivas.
- c) Eliminar los proyectos que no sean particulares.
- d) Recopilar las fojas, número y año de todos los roles anteriormente levantados.

1.4.2 TAREAS ASOCIADAS CON SEGUNDO OBJETIVO

- a) Levantar las transacciones realizadas por cada rol a través de su foja, número y año en la página del Conservador.
- b) Rescatar las fechas, valor y responsables de las transacciones.
- c) Elaborar una nueva base de datos que aglutine los precios anteriormente levantados con las nuevas variables determinadas como relevantes.

1.4.3 TAREAS ASOCIADAS CON TERCER OBJETIVO

- a) Realizar la regresión con la base de datos ya elaborada y construir el modelo que explique del precio de suelo.
- b) Validar el modelo, y que se asemeje a la realidad.
- c) Verificar la veracidad del modelo analizando la heterocedasticidad, autocorrelación y multicolinealidad.

1.4.4 TAREAS ASOCIADAS CON CUARTO OBJETIVO

- a) Interpretar los coeficientes entregados por la regresión lineal.
- b) Dar un sentido a las interpretaciones de los coeficientes.
- c) Dar un enfoque a estos coeficientes para las inmobiliarias.

1.4.5 TAREAS ASOCIADAS CON QUINTO OBJETIVO

- a) Hacer conclusiones generales del estudio realizado
- b) Concluir con una propuesta que continúe con la investigación.

1.5 ALCANCES

En la presente memoria, se lleva a cabo una investigación y estimación del precio del suelo en la comuna de Las Condes. Para ello, se realiza un estudio basado en los datos de los permisos de edificación de proyectos seleccionados en la comuna.

Con el objetivo de recopilar los datos necesarios, se incluyen aquellos proyectos que cumplen con criterios específicos. Estos criterios se basan en la exclusión de transacciones realizadas por particulares en los terrenos, centrándose en proyectos de mayor envergadura y alcance en comparación con otros proyectos.

En el contexto de esta investigación, se entiende por proyectos de mayor envergadura aquellos que involucran construcciones de mayor tamaño y magnitud. Estos proyectos pueden ser llevados a

cabo tanto por empresas como por personas particulares, aunque estos últimos se consideran excepcionales.

Por otro lado, se hace referencia a casos individuales específicos, los cuales se refieren a situaciones en las que se realiza una modificación o ampliación de una edificación ya existente en lugar de una construcción completamente nueva. Estos casos no se consideran relevantes para la base de datos que se está construyendo, ya que no se considera que tengan un impacto significativo en el precio del suelo.

En esta memoria se utiliza como estudio los precios de suelo de la comuna de Las Condes, esta comuna fue elegida por las siguientes razones: a) Ubicación privilegiada: Las Condes es una comuna estratégicamente ubicada en la Región Metropolitana de Santiago, Chile. Se encuentra en una zona de alto desarrollo económico y cuenta con una excelente conectividad vial y de transporte público, lo que la convierte en una ubicación atractiva para vivir y establecer negocios. b) Desarrollo inmobiliario significativo: En Las Condes, se han desarrollado numerosos proyectos inmobiliarios de envergadura, tanto residenciales como comerciales. Esto implica una alta actividad en el mercado inmobiliario y una diversidad de opciones de vivienda, lo que resulta interesante para analizar la dinámica y evolución del precio del suelo. c) Disponibilidad de datos: Al ser una comuna importante y con un mercado inmobiliario activo, es probable que exista una disponibilidad adecuada de datos relacionados con las transacciones de suelo y propiedades en Las Condes. Esto facilita la recopilación y análisis de información necesaria para llevar a cabo un estudio sobre el precio del suelo (Municipalidad de Las Condes, 2022).

2. MARCO TEÓRICO Y REVISIÓN BIBLIOGRÁFICA

En este capítulo, se aborda el marco teórico y la revisión bibliográfica donde se informará acerca de los conceptos más claves a conocer a la hora de abordar el tema de esta memoria, también se profundizará en la historia y acontecimientos claves sobre el precio del suelo en Chile.

2.1 IMPORTANCIA DEL PRECIO DE SUELO

El precio del suelo es de vital importancia en el contexto de la inversión inmobiliaria y en la valoración de bienes raíces en general. Esto se debe a varias razones clave:

- a) **Determina la Rentabilidad de una Inversión:** Para los inversores inmobiliarios, el precio del suelo es un factor crítico que afecta directamente a la rentabilidad de su inversión. Si el precio del suelo es alto, el costo total de adquirir la propiedad será mayor, lo que puede reducir el margen de beneficio. Por lo tanto, conocer el precio del suelo es esencial para evaluar la viabilidad financiera de un proyecto inmobiliario (DÍAZ, 2014).
- b) **Influye en las Decisiones de Desarrollo:** El precio del suelo tiene un impacto significativo en las decisiones de desarrollo. Los inversores y desarrolladores deben decidir qué tipo de proyecto inmobiliario construir en un terreno en función de su costo y de la demanda del mercado (Torrealba, 2019). Por ejemplo, un terreno con un precio elevado puede justificar la construcción de un edificio de departamentos de lujo, mientras que un terreno con precios más bajos puede ser más adecuado para viviendas unifamiliares.
- c) **Determina el Valor de Propiedades Existentes:** El precio del suelo también afecta el valor de las propiedades existentes en una determinada área. Cuando el precio del suelo aumenta en una zona, generalmente impulsa el aumento de los valores de las propiedades en esa área, lo que puede ser beneficioso para los propietarios actuales (Socovesa, 2023).
- d) **Impacta en la Planificación Urbana:** Las autoridades y planificadores urbanos utilizan el precio del suelo como una métrica importante para tomar decisiones sobre el desarrollo y la infraestructura de una ciudad o región. Puede influir en la zonificación de áreas, la inversión en infraestructura pública y la toma de decisiones sobre políticas de desarrollo (Guzman, 2020).

2.2 LA UTILIDAD DEL CÁLCULO DEL PRECIO DEL SUELO PARA LAS EMPRESAS INMOBILIARIAS EN SUS PROYECTOS

Para las inmobiliarias, el cálculo del precio del suelo es una pieza fundamental en la planificación y ejecución de proyectos de desarrollo inmobiliario. Esta información es esencial para tomar decisiones financieras sólidas y estratégicas que afectan directamente la rentabilidad y el éxito de un proyecto.

El precio del suelo influye directamente en el diseño del proyecto. Una ubicación con suelo costoso puede justificar la construcción de propiedades de lujo o edificios de gran altura, mientras que un suelo más asequible puede ser más adecuado para proyectos de viviendas económicas.

Calcular el precio del suelo es esencial para estimar la rentabilidad de un proyecto. Las inmobiliarias evalúan si los costos del suelo, junto con los costos de construcción y otros gastos, se traducirán en ganancias sólidas una vez que las propiedades se vendan o alquilen.

El conocimiento del precio del suelo permite a las inmobiliarias ser competitivas en el mercado. Pueden ajustar sus estrategias de precios para ofrecer propiedades que se ajusten a la demanda y la capacidad adquisitiva de los compradores (Torrealba, 2019).

Con información sobre el precio del suelo en diferentes ubicaciones, las inmobiliarias pueden diversificar su portafolio. Pueden optar por invertir en múltiples proyectos en diferentes áreas para mitigar riesgos y aprovechar oportunidades en el mercado.

Los bancos y otras instituciones financieras a menudo requieren información detallada sobre el precio del suelo como parte de la solicitud de financiamiento para proyectos inmobiliarios. Tener esta información precisa es crucial para obtener respaldo financiero.

El precio del suelo también se utiliza para evaluar los riesgos asociados con un proyecto. Las fluctuaciones en los precios del suelo pueden indicar tendencias en el mercado, lo que ayuda a las inmobiliarias a prepararse para posibles desafíos.

Comprender los factores que influyen en el precio del suelo en una ubicación específica es esencial para tomar decisiones informadas y rentables. Estos factores pueden incluir la demanda del mercado local, la disponibilidad de servicios públicos, la proximidad a infraestructuras clave como transporte y educación, las regulaciones gubernamentales y ambientales, y la situación económica general (finmodelslab, 2023). Al evaluar estos elementos, las inmobiliarias pueden determinar si un proyecto es viable y, de ser así, qué tipo de desarrollo es más adecuado. Esta información también es crucial para establecer precios de venta competitivos y garantizar que el proyecto sea financieramente sólido a largo plazo.

2.3 LA DINÁMICA ENTRE PRECIOS DE SUELO Y VIVIENDA

Para estudiar el precio del suelo, resulta de gran utilidad analizar detenidamente el precio de la vivienda y sus relaciones con diversas variables. Esto se debe a que el precio de la vivienda y el precio del suelo están intrínsecamente vinculados en el mercado inmobiliario.

El precio de la vivienda proporciona una vista panorámica de la demanda y el valor general en una ubicación. Cuando los precios de la vivienda aumentan, a menudo indica un aumento en la demanda de viviendas en esa área. Esto, a su vez, puede influir en el precio del suelo cercano. La relación entre el precio de la vivienda y el precio del suelo es fundamental para entender las oportunidades de inversión, evaluar la rentabilidad de proyectos y tomar decisiones sólidas en el mercado inmobiliario.

Existen estudios en los que se estudió de igual manera que este trabajo, pero como centro de estudio el precio de la vivienda. En ellos se encontró que varios factores son estadísticamente significativos para determinar los precios de las viviendas en Chile, como la antigüedad, la superficie construida y la superficie del terreno. La ubicación también influye, con diferencias en las comunas y el ingreso promedio del vecindario (Sagner, 2011)

2.4 MODELO UTILIZADO PARA PRECIOS DE VIVIENDA

Existen diversos estudios que análogamente a este trabajo, realizó un modelo de regresión lineal con el propósito de buscar que variables explicativas del precio de la vivienda. Estos resultados de las variables pueden ser de gran utilidad para este trabajo para compararlos con los resultados de las regresiones.

El autor Eugenio Figueroa deduce que existen 8 variables que muestran una influencia estadísticamente significativa para el precio de la vivienda. Entre ellas, las que afectan positivamente al valor y las que afectan negativamente, las variables que afectan positivamente son: superficie de terreno, existencias de dependencias de servicio, nivel socioeconómico y densidad de construcción. Por otro lado, las variables que afectan negativamente son las siguiente: superficie edificada, número de habitaciones, pertenencia al área céntrica de la ciudad y densidad poblacional del barrio (Figueroa, 2006).

La ubicación en el área céntrica de la ciudad tiene un fuerte impacto negativo, disminuyendo el valor de mercado en más del 20%. Las características socioeconómicas del vecindario, como los ingresos promedios, tienen un impacto significativo en el precio de las viviendas en Santiago, pudiendo aumentar el valor de una casa en más de ocho veces al cambiar de un vecindario de ingresos bajos a uno de ingresos altos (Figueroa, 2006).

2.5 EL IMPACTO DE LA PLANIFICACIÓN URBANA EN EL PRECIO DEL SUELO

El plan regulador es una variable relevante dado que afectan la accesibilidad, la infraestructura y los servicios de una comuna, lo cual puede influir en la demanda de viviendas y, en consecuencia, en el precio del suelo. Además, promover políticas públicas que reduzcan la brecha entre las comunas y mejoren la calidad de vida de los ciudadanos contribuirá a la construcción de una ciudad mejor y más equitativa (Cifuentes, 2020).

Un plan regulador actualizado es crucial para garantizar un desarrollo urbano ordenado y equilibrado. Establece normativas y pautas para el uso del suelo, la densidad de construcción, las

áreas verdes y otros aspectos urbanísticos. Este puede ayudar a crear comunas atractivas, funcionales y sostenibles, lo que a su vez puede influir en el precio del suelo (Cifuentes, 2020).

En este trabajo las normativas del plan regulador se ven reflejadas en las variables estudiadas, en ellas se refleja el valor o el cumplimiento de la normativa estudiada. Estas variables corresponden a:

- a) Coeficiente de constructibilidad: Dicta el máximo de metros cuadrados a construir en el terreno a través de una fórmula.
- b) Altura máxima: Altura máxima que puede existir entre el suelo natural y el plano paralelo superior.
- c) Rasantes: Recta imaginaria con ángulo determinado por el cual solo se puede construir en ese rango.
- d) Coeficientes de ocupación de suelo: Relación aritmética requerida que existe entre la superficie total construida y la superficie total del terreno.
- e) Estacionamiento requerido: Estacionamiento mínimo a construir dictado por las normativas de la comuna, dependiendo de las características de la edificación.
- f) DFLN2: Norma que dicta disposiciones especiales a todas edificaciones que tienen una superficie edificada y útil de 140 metros cuadrados o menos.
- g) Beneficio de fusión: Norma que dicta si se puede o no fusionar dos o más terrenos
- h) Proyección sombras: Perspectiva de las sombras requerida sobre el terreno.

2.6 REGRESIÓN LINEAL

La regresión lineal, en el contexto de este estudio, representó un modelo estadístico empleado para analizar la relación entre una variable dependiente (precio de suelo) y múltiples variables independientes. Esta técnica permitió evaluar cómo las variables independientes influyeron en la variable dependiente, es decir, cómo ciertas características o factores afectaron a una medida

particular. Cuando tenemos más de una variable independiente, como en este estudio, le llamamos regresión lineal múltiple. (Espinoza, 2017).

Se utilizaron matemáticas y estadísticas para descubrir cuánto impacto tienen estas variables independientes en el precio del suelo. Pero, para hacerlo de manera adecuada, debemos hacer algunas suposiciones específicas y aplicar métodos estadísticos.

La regresión lineal múltiple te entrega como resultado una ecuación de la siguiente manera:

$$Y = \beta_0 + \sum \beta_i X_i + \varepsilon_i$$

Ecuación 1: Ecuación de modelo regresión lineal

Fuente: Creación propia

- a) Y: Variable dependiente resultante de la ecuación.
- b) β_0 : Término independiente, es el valor esperado de Y cuando los X son iguales a 0
- c) X_i : Variables independientes.
- d) β_i : Coeficientes parciales, efecto que tiene el aumento de una unidad de la variable independiente que lo acompaña.
- e) ε_i : Error de la observación provocado por variables no controladas.

De esta manera es como se presenta la ecuación de la regresión lineal múltiple en el que los coeficientes y los errores son entregados por esta para construirla. (Espinoza, 2017).

2.7 REGRESIÓN LINEAL EN RSTUDIO

Una vez finalizada la recopilación de datos, se inicia el proceso de análisis utilizando la plataforma RStudio. RStudio es un software estadístico basado en el lenguaje de programación R, ampliamente utilizado en la investigación y análisis de datos. En este entorno, se llevará a cabo un análisis de regresión lineal con el objetivo de comprender cómo los diversos factores influyen en el valor del suelo (Berrendero, 2016).

El análisis arroja varios resultados importantes:

- a) **Significancia de las variables:** La regresión determina si cada variable independiente tiene un impacto significativo en la variable dependiente. Si una variable no es significativa, se considera que no influye en el valor del suelo y puede eliminarse del modelo.
- b) **Coefficientes de regresión:** Cada variable independiente que se mantiene en el modelo tiene un coeficiente de regresión asociado. Este coeficiente indica cuánto cambia la variable dependiente por unidad de cambio en la variable independiente correspondiente. Es una medida cuantitativa de la influencia de cada factor en el valor del suelo.
- c) **Filtrado de variables explicativas:** Una vez que se ha construido el modelo de regresión, es posible identificar qué variables explican de manera significativa las variaciones en el valor del suelo. Esto permite enfocar el análisis en los factores más relevantes y descartar aquellos que no tienen un impacto significativo.

2.8 CARET Y METODO LVQ

Caret (Classification and Regression Training) es una herramienta integral de RStudio que desempeña un papel crucial en la evaluación y selección de atributos de datos. Su función principal radica en identificar la relevancia y la importancia de las variables que componen un conjunto de datos. No solo proporciona esta información, sino que también facilita la selección de las variables más significativas para un análisis de regresión.

Dentro de las técnicas utilizadas en caret, se encuentra el método Learning Vector Quantization (LVQ). Este es un algoritmo de aprendizaje supervisado con aplicaciones en reconocimiento de patrones y clasificación. LVQ opera en un entorno donde los datos se representan como vectores en un espacio multidimensional. Su objetivo principal es asignar con precisión las clases a estos vectores mediante el entrenamiento de una red neuronal.

Durante el proceso de entrenamiento, la red neuronal ajusta los pesos asociados a cada vector para que se asemejen a los patrones de una clase específica. Esto significa que LVQ ayuda a distinguir entre diferentes categorías al permitir que los vectores de entrada se ajusten de manera óptima a las clases a las que pertenecen. (Guerrero, 2016).

En el contexto de "caret", el método LVQ puede ser utilizado como un algoritmo de clasificación supervisada. Aquí, se pueden aplicar diversas estrategias de validación cruzada y ajuste de parámetros para mejorar el rendimiento del modelo de clasificación. Esta combinación de LVQ con caret permite no solo identificar las variables más importantes, sino también entender cómo contribuyen estas variables a la clasificación en función de su relevancia.

Caret utiliza diferentes métodos y técnicas para determinar la relevancia e importancia de las variables en un modelo de clasificación o regresión. Algunas de las estrategias clave que caret emplea para esta selección son:

- a) Recursive Feature Elimination (RFE): Este método implica entrenar el modelo inicial con todas las variables y luego eliminar iterativamente las menos importantes. Esto se hace evaluando el rendimiento del modelo después de cada eliminación. Las variables que contribuyen menos a la precisión del modelo se eliminan hasta que se alcanza un conjunto óptimo de características (Kuhn, 2019).
- b) Importancia de variable: Caret utiliza algoritmos específicos, como Random Forest y Gradient Boosting, para calcular la importancia relativa de cada variable. Estos algoritmos evalúan cómo cada variable afecta la precisión del modelo y generan una puntuación de importancia para cada una (Kuhn, 2019).
- c) Correlación entre Variables: Caret también considera la correlación entre las variables. Si dos variables están altamente correlacionadas (es decir, proporcionan información similar), es posible que solo se seleccione una de ellas para evitar la multicolinealidad y mejorar la interpretabilidad del modelo.
- d) Validación cruzada: La validación cruzada se utiliza para evaluar el rendimiento del modelo con diferentes conjuntos de datos de entrenamiento y prueba. Durante este proceso, caret puede observar cómo el rendimiento del modelo varía al agregar o eliminar ciertas variables. Esto proporciona información sobre la importancia de las variables en diferentes contextos (Kuhn, 2019).
- e) Métodos Específicos del Modelo: Dependiendo del algoritmo de clasificación o regresión utilizada, caret puede aplicar métodos específicos para seleccionar características (Kuhn, 2019).

Además, es importante mencionar que caret no solo se enfoca en la relevancia de las variables, sino que también evalúa la correlación entre ellas. Esto significa que las variables altamente correlacionadas pueden recibir menos importancia en el proceso de selección, ya que aportan información similar.

2.9 HETEROCEDASTICIDAD

La heterocedasticidad es una característica de los datos que se produce cuando la varianza de los errores de un modelo estadístico no es constante a lo largo de las observaciones. En otras palabras, la heterocedasticidad significa que los errores del modelo tienen una varianza diferente para diferentes niveles de la variable explicativa o predictor.

La heterocedasticidad es importante de detectar porque puede tener varios efectos en los resultados del modelo. En particular, puede afectar la precisión de las estimaciones y las inferencias estadísticas. Cuando los errores tienen una varianza diferente en diferentes niveles de la variable explicativa, los errores pueden estar sesgados hacia un lado, lo que puede afectar la eficiencia y la precisión de las estimaciones de los parámetros del modelo y una disminución en la precisión de las estimaciones (Jara, 2008).

En este trabajo se realizó una prueba llamada Breusch-Pagan en el programa de RStudio, este es una prueba estadística utilizada en econometría para evaluar si un modelo de regresión lineal muestra evidencia de heterocedasticidad (Flores, 2016). La prueba de Breusch-Pagan funciona mediante la comparación de la suma de cuadrados de los residuos (SSR) de un modelo de regresión lineal con el mismo modelo de regresión, pero con una variable adicional que representa la variable explicativa multiplicada por los residuos al cuadrado. Si la suma de cuadrados de esta variable adicional es significativa, entonces se concluye que hay heterocedasticidad en los errores del modelo. En esta prueba la hipótesis nula es que los errores de un modelo de regresión lineal tienen varianza constante (es decir, no hay heterocedasticidad). La hipótesis alternativa es que los errores tienen varianza, es decir, hay heterocedasticidad. Para evaluar la hipótesis nula, se usa una prueba

F, y si el valor p asociado con la prueba F es menor a 0.05, se rechaza la hipótesis nula y se concluye que hay evidencia de heterocedasticidad.

2.10 MULTICOLINEALIDAD

La multicolinealidad es un fenómeno en el cual dos o más variables independientes en un modelo estadístico están altamente correlacionadas entre sí. Esto puede conducir a una estimación imprecisa de los coeficientes y una disminución de la capacidad del modelo para predecir los valores de la variable dependiente. (Asturias Corporación Universitaria, s.f.).

Para medir la multicolinealidad se puede utilizar el factor de inflación de la varianza (VIF), es una medida utilizada en el análisis de regresión para evaluar la multicolinealidad entre las variables independientes. La multicolinealidad se refiere a la alta correlación o dependencia lineal entre dos o más variables independientes en un modelo de regresión.

El VIF se calcula para cada variable independiente en el modelo y cuantifica cuánto aumenta la varianza de un coeficiente de regresión debido a la multicolinealidad con otras variables. Matemáticamente, el VIF se calcula como el cociente de la varianza del coeficiente estimado de una variable dividida por la varianza del coeficiente estimado de esa misma variable cuando se ajusta en un modelo de regresión lineal múltiple (CFI , 2022).

El VIF se interpreta de la siguiente manera:

1. Un VIF igual a 1 indica que no hay multicolinealidad entre la variable independiente y las demás variables del modelo. No hay correlación lineal alta.
2. Un VIF mayor que 1 e inferior a 5 generalmente se considera aceptable, lo que indica una correlación moderada entre la variable independiente y las demás variables del modelo.
3. Un VIF mayor que 5 indica una alta multicolinealidad, lo que implica una correlación fuerte entre la variable independiente y las demás variables del modelo. En este caso, se debe considerar la eliminación o transformación de las variables con alto VIF.

2.11 AUTOCORRELACIÓN

La autocorrelación es una violación de la suposición de independencia de errores en un modelo estadístico, lo que significa que los errores están correlacionados entre sí en lugar de ser independientes. Esto puede conducir a una sobrestimación de la significancia estadística de los coeficientes, lo que lleva a la selección incorrecta de variables independientes y modelos subóptimos. (Asturias Corporación Universitaria, s.f.).

Para medir la correlación se puede utilizar la prueba de Durbin-Watson, esta es una prueba estadística utilizada para detectar la presencia de autocorrelación en los residuos de un modelo de regresión.

La prueba de Durbin-Watson calcula un estadístico que varía entre 0 y 4. Un valor de 2 indica que no hay autocorrelación en los residuos, mientras que valores cercanos a 0 indican autocorrelación positiva (los residuos están correlacionados positivamente entre sí), y valores cercanos a 4 indican autocorrelación negativa (Ramirez, s.f.).

La prueba de Durbin-Watson se utiliza principalmente en modelos de regresión lineal para verificar la suposición de independencia de los errores o residuos. Si se encuentra evidencia de autocorrelación en los residuos, significa que el modelo de regresión no captura completamente la estructura de dependencia presente en los datos. En tales casos, se requieren técnicas o ajustes adicionales para tener en cuenta la autocorrelación y obtener estimaciones y pruebas de hipótesis confiables.

3. LEVANTAMIENTO DE DATOS Y GENERACIÓN DE BASE DE DATOS

En este capítulo, se describe la metodología utilizada para el levantamiento de datos, se expone las selecciones debidas y cómo se filtró para generar la base de datos. Además, se da a conocer la forma en que se agrupan los datos y qué variables afectan al precio de suelo.

3.1 LEVANTAMIENTO DE DATOS

En este inciso, se redacta cómo fue el proceso de levantamiento de datos y todos los pasos que se siguieron para conseguir la información necesaria y relevante con respecto al precio del suelo.

3.1.1 LEVANTAMIENTO DE PERMISOS E IDENTIFICACIÓN DE PROYECTOS

Con el propósito de analizar y conocer de mejor manera las variables para construir el modelo explicativo del precio de suelo, es necesario contar con una base de datos robusta que sea significativa para la comuna de Las Condes. Con este fin, se levantaron los permisos de edificación más representativos. Sin embargo, surge la pregunta ¿qué variables son consideradas significativas?

En respuesta a la pregunta anterior se debe considerar lo siguiente: a) tamaño de la edificación y quién es su propietario. Este último puede ser una persona o una razón social. No obstante, en este trabajo solo se incluyen edificaciones de empresas y excepcionalmente personas particulares. b) se debe clasificar si en el terreno se está edificando algo nuevo o se está modificando/ampliando alguna edificación ya construida. El segundo caso no es de utilidad para la base de datos, ya que una ampliación en algún terreno no es significativa para el precio de la vivienda. c) otro factor que se debe tener en cuenta corresponde a la dimensión del edificio que se construye. En este caso particular, se tomaron en cuenta proyectos de 3 o más viviendas o simplemente edificios.

Los factores mencionados anteriormente se deben tener en cuenta por varias razones: a) por la representatividad de la muestra, si se incluyen proyectos pequeños es posible que se sesgue la representatividad y se obtengan conclusiones que no sean generalizables para el conjunto de la población de edificaciones en la ciudad. b) eficiencia en la recolección de datos, incluir estos

proyectos en la muestra puede requerir un esfuerzo adicional para recopilar la información necesaria. c) significancia estadística, puede aumentar la variabilidad de los datos y disminuir el poder estadístico del análisis (Novales, 2010).

Todos estos permisos se recopilaron a través de la Página de Transparencia de la comuna de Las Condes desde 2012 hasta el 2021.

En este trabajo se analizó un total de 1542 permisos de edificación, y se determinó si cada uno de ellos era o no relevante. En virtud de esto, solo se incluyeron 455 permisos de edificación con sus respectivos datos.

Tabla 1: Permisos de edificación levantados

Numero	Fecha	Fecha de publicación	Breve descripción	Enlace de publicación	Permiso utilizable
176	22-12-2021	12-01-2021	Condominio tipo A de 4 viviendas	pdf	Si
169	07-12-2021	12-01-2021	Parroquia Nuestra Señora de los ángeles	pdf	Si
168	07-12-2021	12-01-2021	Multicancha Techada Colegio Acogido	pdf	Si
167	06-12-2021	12-01-2021	Conjunto de 19 viviendas unifamiliar	pdf	Si
162	23-11-2021	10-12-2021	Condominio de 23 viviendas unifamiliares en la quebrada	pdf	Si
161	23-11-2021	10-12-2021	Condominio tipo A de 26 viviendas	pdf	Si
159	19-11-2021	10-12-2021	Condominio tipo A de 12 viviendas	pdf	Si
157	17-11-2021	10-12-2021	Un edificio de 7 pisos + 5 subterráneos	pdf	Si
155	12-11-2021	10-12-2021	Cinco viviendas de 2 pisos + subterráneo	pdf	Si

Fuente: Elaboración propia

En la tabla 1, se presentan los permisos de edificación levantados en los que gracias a “breve descripción”, es posible clasificar si el permiso es utilizable o no. Esta base de datos, en enlace a

la publicación, entrega el archivo que contiene el permiso mismo. Esta tabla es una muestra de la tabla de Excel¹.

3.1.2 IDENTIFICACIÓN DE ROLES

En los permisos significativos, se levantó el número de manzana, de predio, tipo de norma, número de la norma, fecha del permiso, fecha de publicación, descripción del objeto del acto y el número de aprobación. Todos estos datos se emplearán, posteriormente, en el escrito. Con los números de manzana y predio, se puede obtener mucha información acerca de los terrenos, ya que estos números identifican el terreno delimitado específicamente. Un permiso de edificación puede tener uno o más números de manzana y predio, por lo tanto, los datos aumentan significativamente en comparación con la cantidad de permisos. Cabe destacar que también existe un número de comuna que se refiere a aquella en la cual está ubicada, pero como en este levantamiento solo se considera la comuna de Las Condes, no se utiliza este antecedente para la base de datos. Cada combinación de predio, con su debida manzana, significa varias transacciones de los terrenos a lo largo de los años, según el número de veces que se hizo una transacción por el terreno. Estas transacciones se identifican con tres variables: foja, número y año.

En este trabajo, del total de 544 permisos levantados, se le identificaron sus roles, por lo que los datos aumentaron consideradamente a 1.534, debido a la cantidad de roles de predio que existía en cada permiso de edificación. Esto generó una base de datos más amplia en la que se rescata el valor de cada uno de estos roles y se clasifica el terreno con mayor propiedad (Sii, 2022).

¹ Tabla en Anexo 1

Tabla 2: Roles identificados

Numero	FECHA	Cod mz	Cod pr
176	22-12-2021	2640	49
169	07-12-2021	245	2
168	07-12-2021	2600	20
167	06-12-2021	1204	197
162	23-11-2021	1204	194
161	23-11-2021	1204	201
159	19-11-2021	2164	6
157	17-11-2021	1597	10
157	17-11-2021	1597	11
157	17-11-2021	1597	12

Fuente: Elaboración propia

En la tabla 2, se presenta una muestra parcial de una de las bases de datos de los roles en la que se asocia cada proyecto con cada uno de sus roles respectivos. En este caso, la cantidad de roles de un proyecto varía según cada caso. En la tabla se representa con número a los números del permiso de edificación, Cod mz que corresponde al código de manzana correspondiente y Cod pr que es el código de predio del terreno.

3.1.3 LEVANTAMIENTO DE TRANSACCIONES

Por medio de los tres valores, se puede identificar el documento de la compra por el terreno. Con el fin de levantar estas tres variables, se debe identificar al número de manzana y predio al cual están referenciados. Posteriormente, con el número de manzana y predio se puede identificar el permiso de edificación y con este el nombre del propietario del terreno y del año en que se realizó el permiso. Con esta información, a través de la página del conservador.cl, se puede identificar de dos formas el documento de compra por un terreno: una forma es con el nombre del propietario, comuna y año de la transacción. Gracias a los datos del paso anterior, se pueden levantar la foja, el número y año de la transacción del predio. Al ingresar estos datos a la página, se obtienen todas las transacciones del terreno a lo largo de los años. En este caso, solo son útiles los documentos a nombre de los propietarios del terreno con la fecha anterior al de la edificación. No son relevantes los documentos de transacción posteriores a esta, ya que si se utilizan transacciones posteriores a la fecha de edificación no serán relevantes con los datos respectivos del permiso.

En esta extensa base de datos, solo se utilizaron los documentos relacionados con los suelos edificados entre los años 2017 y 2021. De esta forma, se reduce la cantidad de datos de 1.534 a 638, lo que constituye una base de datos significativa para la construcción del modelo.

Inicialmente, se consideró aplicar el método de web scraping, que consiste en extraer datos automáticamente de sitios web, para obtener un conjunto más amplio de transacciones de terrenos. Sin embargo, se encontró un obstáculo importante: los documentos de las transacciones de terrenos están en formato escaneado, lo que impide el reconocimiento automático por parte de las técnicas de web scraping.

Además, es importante destacar que las consultas en la página web estaban restringidas a un límite de 15 por día, lo que hacía que el proceso de recopilación de datos fuera extremadamente lento y poco práctico. Como consecuencia de estas limitaciones técnicas y operativas, se tomó la decisión de trabajar únicamente con la cantidad de documentos obtenidos de manera manual.

Tabla 3: Documentos de transacción

Código Manzana	Código predio	Fojas	Numero	Año	Comprador	Vendedor	Fecha inscripción	Precio UF	Precio CLP	Fecha escritura
2640	49	43816	64097	2021	Inmobiliaria la fuente	Ruiz Luci Raúl	15-06-2021	36500		27-04-2021
1204	197	51615	75373	2021	Inmobiliaria Lima de Asís	La Fundación Cultural Las Condes	15-07-2021	114314		29-06-2021
1204	194	51614	75372	2021	Inmobiliaria Lima de Asís	La Fundación Cultural Las Condes	15-07-2021	138379		29-06-2021
2164	6	9260	13645	2021	Reihenhaus	Telias Ergas	20-02-2021		363.695.125	21-01-2021

Fuente: Elaboración propia

En la Tabla 3 se observa una muestra de los datos levantados de los documentos de transacción². Estos corresponden a: a) código de manzana y de predio, para identificar el predio del terreno. b) foja, número y año del terreno correspondientes a los códigos anteriores. c) nombres de compradores y vendedores del terreno. d) fechas de inscripción y escritura del documento. e)

² Tabla de Excel en Anexo 2

Precio por el que se tranzó el terreno, este puede estar en UF o en pesos chilenos, dependiendo del que se utilizó en el documento.

Una vez levantados todos los valores de los terrenos, se agruparon los predios por sus respectivos sectores que se tranzaron en el permiso de edificación, ya que en este se encuentra el valor total del terreno en metros cuadrados. Posteriormente, se creó una nueva variable en la cual se dividió el valor total por el que se tranzó el terreno, por la cantidad de metros cuadrados. De esta forma, se obtuvo el valor de cada metro cuadrado por proyecto. Esta base de datos entregó un total de 200 edificaciones, ya que los predios se unieron por permiso de edificación. De esta nueva variable, se levantaron los siguientes datos.

Tabla 4: Estadísticas valor unitario en UF

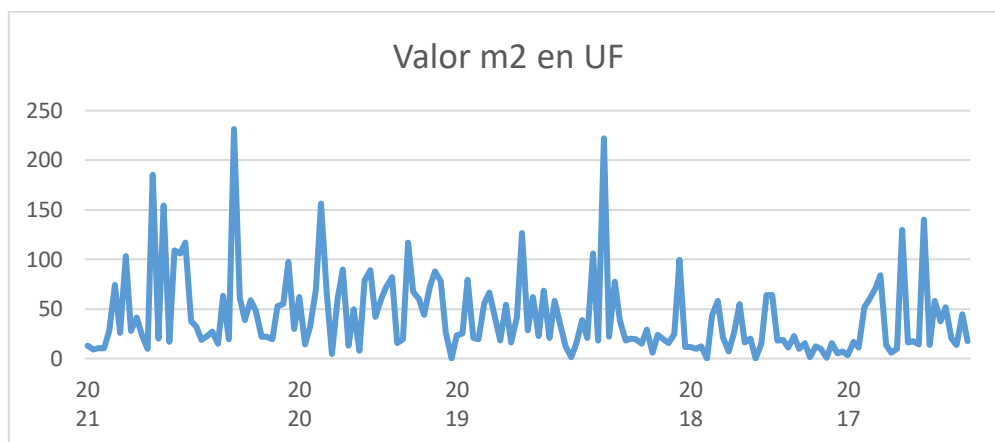
Media valor m2 2017	Media valor m2 2018	Media valor m2 2019	Media valor m2 2020	Media valor m2 2021	Media valor m2 total
29.11108	17.35448	34.43006	48.16027	45.7395	35.1614
Desviación Estándar 2017	Desviación Estándar 2018	Desviación Estándar 2019	Desviación Estándar 2020	Desviación Estándar 2021	Desviación Estándar total
36.79236	18.71874	39.54556	38.85458	50.81069	40.37743

Fuente: Elaboración propia

Como se puede observar en la Tabla 4 las medias de los valores por metro cuadrado de terreno varían considerablemente en cada año, y no necesariamente siguen un patrón a medida que pasan los años, esto se debe a la variabilidad de los terrenos, los terrenos tienen muchos factores que influyen en su precio, incluyéndose en estos factores el año. Anticipadamente, se puede decir que la variable año es significativa por su variabilidad con los años, pero se debe realizar un análisis más profundo como el que se hará más adelante en el escrito. También se pueden observar desviaciones estándar muy cercanas a la media, lo que implica que los datos tienen una dispersión

considerable alrededor de la media. La mayoría de los datos estarán a una distancia considerable de la media, ya sea por encima o por debajo de ella (Carmona, s.f.).

Ilustración 1: Gráfico valor m2 en UF



Fuente: Elaboración propia

En la Ilustración 1 se puede respaldar lo dicho anteriormente en que los valores de cada metro cuadrado varían considerablemente uno con otro.

3.2 ANÁLISIS DE DATOS

3.2.1 CREACIÓN BASE DE DATOS

Una vez identificado el documento se comienza a trabajar con él, el documento es un informe escrito en el que se identifica las limitaciones del terreno, los miembros que están realizando la transacción junto con su RUT (donde puede ser una o más entidades las que vendan el terreno), las fechas de transacción y de escritura del documento y los montos por el cual fue tranzado. Este monto puede estar en unidades de fomento (UF) o en pesos chilenos. En este caso el valor levantado es en unidades de fomento, para pasar de pesos chilenos a UF solamente se debe buscar el monto de la UF en pesos chilenos en aquel entonces en la página de servicios de impuestos internos, con la fecha de escritura del documento. Todas estas variables anteriormente definidas fueron levantadas para la creación de la base de datos sobre los valores por los que el terreno fue tranzado en su totalidad.

Una vez obtenida esta base de datos se comenzó a trabajar en una nueva, esta es la de los datos de los permisos de la obra, en la que a través de los permisos de edificación se pudo levantar más datos sobre la edificación del terreno. Los datos levantados fueron los siguientes: número de edificaciones, superficie edificada, número de pisos, zona de edificación y uso, uso que se le dará a la edificación, superficie edificada útil, común y total del terreno, coeficiente de constructibilidad, altura máxima, rasantes, coeficiente de ocupación de suelo, densidad de vivienda, adosamiento, antejardín, estacionamiento requerido, DFLN2, conjunto armónico copropiedad inmobiliaria, beneficio de fusión, proyección sombras y las cantidades de viviendas, estacionamientos vehículos, estacionamiento bicicletas y bodegas.

Todas las variables anteriormente mencionadas son variables que influyen en el precio por el que se tranzó la edificación e influyen finalmente en el precio del suelo. Existe una diferencia en cómo afecta cada una de estas variables en el precio, es por esto por lo que con la base de datos generada se realizó una regresión en la que se quiere definir cuan influyentes son cada una de estas variables. A partir de ahí se puede analizar los datos obtenidos y definir porque es que cada variable afecta al precio final.

3.2.2 VARIABLES INDEPENDIENTES

Como se mencionó anteriormente, existen variadas variables independientes que afectan en el precio de suelo, todas estas variables fueron levantadas para la construcción del modelo, a continuación, se listarán y explicarán estas variables dictando también una predicción de su significancia según lo estudiado anteriormente.

- a) Número de edificaciones: El número de edificaciones como dice su nombre hace referencia a la cantidad de edificaciones por terreno.
- b) Superficie edificada: Superficie donde se edificó.
- c) Número de pisos: Cantidad de pisos por edificación.
- d) Zona de edificación y de uso: Hace referencia al tipo de suelo en el que se encuentra la edificación, pueden ser UV0, UV1, UV2, UV3, UC1, UC2 en cuanto a uso y EAa1, EAa2,

EAA3, EAA4, EAb1, EAb2, EAb3, EAb4, Eam1, Eam2, Eam3, Eam4, Ee1, Ee2 y Ee3 en cuanto a edificación.

- e) Uso de la edificación: Finalidad que se le dará a la edificación, pueden ser residencial, hotel, equipamiento, comercio, oficina, actividades productivas e infraestructuras.
- f) Superficie edificada útil: Es toda superficie que no incluya muros, pilares y elementos no prescindibles.
- g) Superficie edificada común: Superficie que separa la edificación útil.
- h) Coeficiente de constructibilidad: Dicta el máximo de metros cuadrados a construir en el terreno a través de una fórmula.
- i) Altura máxima: Altura máxima que puede existir entre el suelo natural y un plano paralelo superior al mismo.
- j) Rasantes: Recta imaginaria con ángulo determinado por el cual solo se puede construir en ese rango.
- k) Coeficientes de ocupación de suelo: Relación aritmética que existe entre la superficie total construida y la superficie total del terreno.
- l) Densidad de la vivienda: Relación entre el número de viviendas del terreno y su superficie.
- m) Adosamiento: Edificación no subterránea que se ubica contigua a los deslindes.
- n) Antejardín: Distancia que existe entre la línea de edificación y la línea oficial del terreno.
- o) Estacionamiento requerido: Estacionamiento mínimo a construir dictado por las normativas de la comuna, dependiendo de las características de la edificación.
- p) DFLN2: Norma que dicta disposiciones especiales a todas edificaciones que tienen una superficie edificada y útil de 140 metros cuadrados o menos.
- q) Conjunto armónico copropiedad inmobiliaria: Hace referencia a aquellos proyectos cuya edificación corresponda a equipamiento comunitario.
- r) Beneficio de fusión: Beneficio a fusionar dos o más terrenos en el que aumenta el coeficiente de constructibilidad del predio.
- s) Proyección sombras: Perspectiva de las sombras sobre el terreno.
- t) Cantidades de destinos a la edificación: Cantidad de roles que se le da a la edificación, estos pueden ser; viviendas, locales comerciales, oficinas, estacionamientos, estacionamiento de bicicleta, bodegas y otros (especificado en cada caso).
- u) Ubicación: Variable compuesta por dos valores: distrito censal y zona

4 TRANSFORMACIÓN Y ANÁLISIS DE RESULTADOS

En este capítulo se redactará todos los procesos relacionados con la regresión lineal que se llevaron a cabo para lograr la regresión más significativa para el precio del suelo, se mencionarán los ajustes a estas variables y como se tomó la decisión de ejecutar estos cambios. Finalmente, se enseñará la regresión final de este trabajo y los análisis respectivos de los resultados entregados por esta.

4.1 TRANSFORMACIÓN VARIABLES

Antes de realizar un análisis de regresión lineal, es importante especificar el modelo y preparar las variables. La especificación del modelo implica seleccionar las variables independientes que deben incluirse y cuáles no, así como analizar si es necesario modificar las variables independientes (Espinoza, 2018).

En el caso de las variables numéricas, se pueden considerar diferentes transformaciones, como la transformación cuadrática y la transformación logarítmica. Sin embargo, es crucial destacar que estas transformaciones no se utilizan necesariamente para normalizar la distribución de los datos, sino más bien para capturar relaciones no lineales o potenciales entre las variables independientes y la variable dependiente.

La transformación logarítmica se utiliza comúnmente cuando la variable dependiente está sesgada hacia la derecha, es decir, cuando la media es mayor que la mediana. Esta transformación puede ayudar a estabilizar la varianza y mejorar la linealidad en los modelos de regresión (Ciencia sin Sesos, 2014).

Por otro lado, la transformación cuadrática se aplica para explorar relaciones no lineales entre la variable dependiente y las variables independientes. Se utiliza cuando se sospecha que la relación entre las variables no puede ser capturada de manera adecuada por un modelo lineal (Ciencia sin Sesos, 2014).

En este trabajo, se realizaron transformaciones a la variable dependiente (precio por metro cuadrado) utilizando solamente la transformación logarítmica. Se ejecutaron histogramas para evaluar los diferentes estados de la variable.

Tabla 5: Ejemplo de transformaciones.

Valor uf m2	Log valor uf m2
-------------	-----------------

13.3	2.58
9.4	2.24
10.1	2.32
10.2	2.32
27.6	3.32
78.2	4.36

Fuente: Elaboración propia.

En la Tabla 5 se puede observar un ejemplo de 6 distintos valores a la variable dependiente y las transformaciones debidas. Esta tabla ayuda a entender cómo se comportan los datos una vez fueron transformados a logaritmo con valores reales de suelo.

Tabla 6: Resumen variable y sus transformaciones.

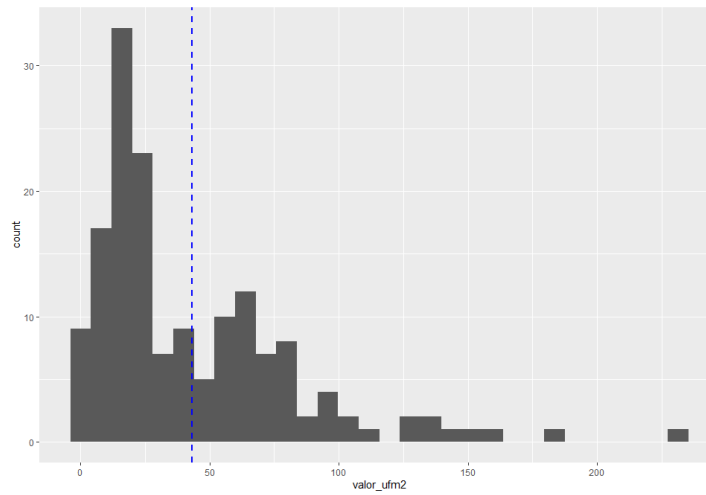
	Valor uf m2	Log valor uf m2
Valor mínimo	0.03831	-3.262
Primer cuartil	15.85287	2.763
Mediana	26.77974	3.287
Media	43.016	3.271
Tercer cuartil	61.73	4.123
Valor máximo	231.427	5.444

Fuente: Elaboración propia.

En la Tabla 6, se puede observar un resumen de los distintos estados de la variable dependiente. Cabe destacar como varía la diferencia entre la media y la mediana en las transformaciones, en el caso de la transformación logarítmica la mediana supera a la media, caso contrario de la variable sin transformación.

En la Ilustración 2, se presenta el histograma de los datos realizado a partir de la variable dependiente sin transformación.

Ilustración 2: Histograma valor metro cuadrado.

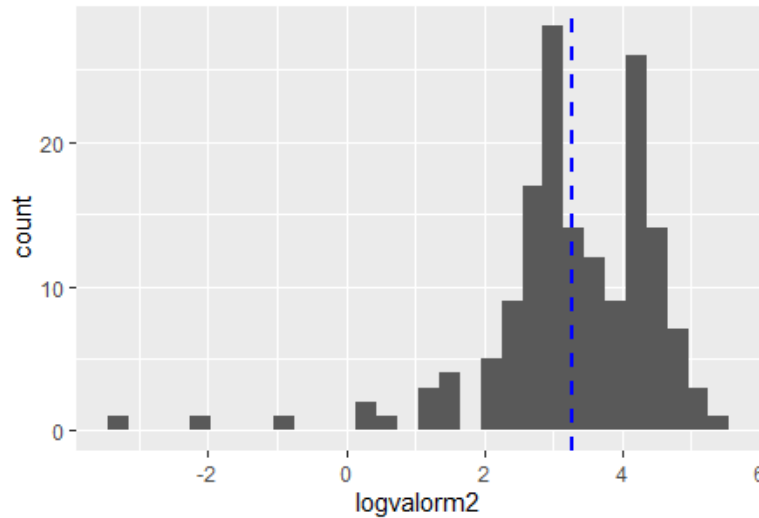


Fuente: Elaboración propia

En la Ilustración 2 se puede observar el histograma realizado a partir de la variable dependiente sin transformar, en el que se puede asumir que no existe normalización de datos. Esto debido a que se observa una concentración de los datos hacia el valor mínimo, sesgado hacia la derecha y, como se plasmó anteriormente, la mediana es considerablemente menor a la media (línea azul).

En la Ilustración 3 se presenta el histograma de los datos ejecutado a partir de la variable dependiente transformada a logaritmo.

Ilustración 3: Histograma variable dependiente en logaritmo.



Fuente: Elaboración propia.

En la Ilustración 3 se puede observar el histograma realizado a partir de la transformación logarítmica a la variable dependiente, en ella se puede deducir una distribución que se aproxime más a la normalidad (debido a que la media se encuentra concentrada en el medio de los datos). Sin embargo, esto no garantiza que la variable dependiente siga una distribución normal.

En esta regresión lineal, se asume la normalidad de los errores o residuos del modelo. Los errores representan la diferencia entre los valores observados y los valores predichos por el modelo. La normalidad de los errores implica que siguen una distribución normal con una media de cero y una varianza constante. Esta normalidad es asumida debido a que se presenta una muestra de tamaño reducido, en caso contrario se asumiría normalidad asintótica (Vela, 2010).

Basándonos en las 2 ilustraciones mencionadas, se eligió la Ilustración 3, que corresponde al histograma de la variable dependiente transformada a logaritmo. La elección se hizo debido a que esta transformación proporciona una distribución que se aproxima más a la normalidad en comparación con las otras opciones.

La Ilustración 2, la cual representa el histograma de la variable dependiente sin transformar, muestra una concentración de datos hacia el valor mínimo y un sesgo hacia la derecha. En contraste, en la Ilustración 3 se muestra el histograma de la variable dependiente transformada a

logaritmo. En esta transformación, se puede apreciar una distribución que se acerca más a una distribución normal. La media de los datos está concentrada en el centro de la distribución, lo cual es deseable para cumplir con uno de los supuestos de la regresión lineal.

Sin embargo, es importante destacar que la normalidad de los errores o residuos en la regresión lineal se asume, no se garantiza, incluso cuando se elige una transformación que aproxima la distribución de la variable dependiente a una distribución normal. La normalidad de los errores es necesaria para cumplir con los supuestos del modelo de regresión lineal y obtener resultados confiables.

Una vez escogido el estado de la variable dependiente se debe escoger el estado de las variables independientes según la dependiente. Las variables transformadas fueron las siguientes: año, superficie edificada, número de pisos, superficie edificada útil, superficie de terreno, coeficiente de constructibilidad, altura máxima, rasantes y coeficiente de ocupación de suelo. El método para escoger el estado de estas variables fue a través de un paquete de RStudio llamado Caret que utiliza el método LVQ mencionado anteriormente, este paquete también te ayuda a elegir con qué variables te quedas y con cuáles no para la regresión. En este paquete se evalúan la importancia relativa de cada variable independiente mediante medidas como la ganancia de información, la importancia de Gini o la reducción de error cuadrático. Ejemplos de estos métodos son Random Forest, Gradient Boosting y árboles de decisión (Amat, 2018). Estas funciones se emplearon para ver la importancia de todas las variables independientes juntos con sus transformaciones, por lo que se eligió la variable con más importancia para la regresión. En el Anexo 3 se puede observar los resultados de caret.

En el caso de la variable de la ubicación, se separó la comuna en sus 16 distritos censales y cada distrito se separó en sus zonas, que su número depende de cada distrito. Después se creó la variable compuesta por dos números; el primero es el valor de la zona en la que se encuentra el terreno y el segundo corresponde al distrito censal correspondido, a este se le llamo zona censal, un ejemplo de esta variable es: Zona censal 3_6.

Después, con ayuda de caret, se eligieron las zonas censales con mayor importancia y se transformaron en variables binarias, en las que si el terreno corresponde a la zona censal transformada en variable se retorna un 1, caso contrario es 0. Esto se hizo debido a que en caret, las dummy, las retorna como la importancia de cada dato por separado, por lo que se decidió convertirlas en binarias, de manera que en la regresión se elijan zonas censales relevantes para la variable dependiente³.

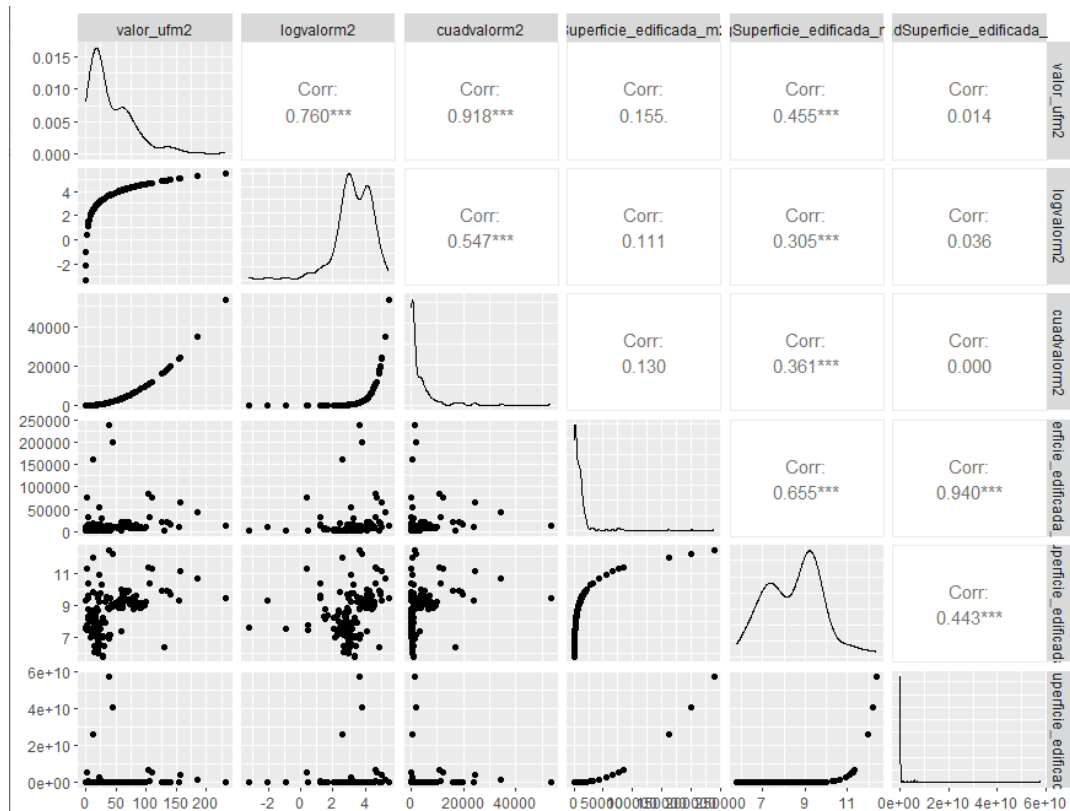
4.2 CORRELACIÓN DATOS

Una vez seleccionadas las variables más importantes para la regresión lineal múltiple a través de las funciones de caret (coeficiente de constructibilidad, número de pisos, rasantes, superficie edificada útil total, altura máxima, año, estacionamientos, superficie edificada, superficie de terreno), se calculó, a modo de análisis, la correlación de los estados de las variables seleccionadas con los estados de la variable dependiente. La correlación existe entre dos variables cuando una de estas otorga información acerca de la otra, esto quiere decir, que afectan a la variable dependiente de igual manera proporcionalmente. Para ver la correlación en los datos mencionados anteriormente se creó la matriz de correlación a través de la función ggpairs de RStudio.

La Ilustración 4 muestra la correlación de los estados de la variable dependiente con los estados de la variable independiente superficie edificada. Cabe mencionar que solamente se están analizando 2 variables.

³ Resultados en Anexo 7

Ilustración 4: Matriz de correlación superficie edificada.



Fuente: Elaboración propia

Las variables comparadas en la matriz de correlación siguen el siguiente orden: valor unitario del suelo, logaritmo del valor unitario, valor unitario al cuadrado, superficie edificada, logaritmo de superficie edificada y superficie edificada al cuadrado. En esta matriz ha de tener en cuenta los valores de correlación. Una correlación alta es cercana tanto a -1 como a 1, por lo que valores altos positivos como negativos son los que se están buscando (Ramon, s.f.). Al analizar el valor de correlación de la variable independiente, se puede observar un valor de 0.036 en el estado cuadrático de la variable con la variable dependiente transformada a logaritmo, lo que es un valor bajo, asimismo ocurre con el estado sin transformar de la variable en la que se obtuvo una correlación de 0.111. En cambio, el estado de logaritmo para la variable independiente tiene una correlación considerable en comparación con el resto con un valor de 0.305. Los gráficos representados en la Ilustración 4 muestran cómo se comportan los datos según las variables que lo conforman, estas variables son las que se interceptan en el eje y con el eje x de la ilustración.

Cabe destacar que el criterio de correlación no es una forma adecuada de seleccionar la forma funcional de la regresión. Para determinar la forma funcional más apropiada, se utilizan medidas de ajuste y criterios de información de modelos. La correlación entre variables solo indica la fuerza y dirección de la relación lineal entre ellas, pero no proporciona información completa sobre la forma funcional óptima del modelo.

Esta matriz se construyó para las siguientes variables: años, superficie edificada, número de pisos, coeficiente de constructibilidad, superficie de terreno, rasantes y coeficiente de ocupación de suelo. Cabe señalar que en muchos casos los valores de correlación son cercanos entre sus estados.

4.3 AJUSTE REGRESIÓN LINEAL MÚLTIPLE

Una vez terminado de estudiar las variables se comenzó a trabajar con la regresión lineal múltiple. Como se dijo anteriormente, esta regresión fue realizada a través del programa RStudio con la librería estándar del programa. A modo de análisis, se realizó la regresión con las variables anteriormente rescatadas de las funciones de caret junto con los estados de las variables consideradas con la matriz de correlación para posteriormente ir modificándola hasta llegar a la regresión más exacta y significativa para la variable dependiente. Las variables seleccionadas para la regresión fueron las siguientes: coeficiente de constructibilidad, año de transacción, número de pisos, densidad, estacionamientos, superficie edificada, superficie de terreno, rasantes, superficie edificada útil, altura máxima, viviendas, número de edificaciones, coeficiente ocupación de suelo, oficinas, locales comerciales y las zonas censales seleccionadas anteriormente.

Primero se mostrará los resultados de la regresión con las variables dichas explicando los resultados de esta para después ver la evolución de esta con los cambios respectivos.

Tabla 7: Primera regresión lineal múltiple

Coeficientes	Estimado	Error Estándar	T value	P value
Intercept	-1.875e2	5.852e1	-3.205	0.001681**

Cuad coef de constructibilidad	8.733e-2	2.671e-2	3.269	0.001362**
Cuad número de pisos	8.703e-4	7.544e-4	1.154	0.250626
Años	9.348e-2	2.899e-2	3.225	0.001575**
Cuad superficie edificada	-3.318e-12	5.353e-11	-0.062	0.950666
Densidad	2.832e-3	2.087e-3	1.357	0.177062
Estacionamientos	-1.343e-3	1.13e-3	-1.188	0.236796
Cuad superficie de terreno	-1.336e-10	7.731e-11	-1.728	0.086302*
Rasantes	4.408e-2	1.367e-2	3.224	0.001576**
Superficie edificación útil total	-4.194e-1	4.853e-1	-0.863	0.388978
Cuad altura máxima	-6.415e-5	2.202e-4	-0.291	0.771279
Numero de edificaciones	7.937e-3	2.083e-2	0.381	0.703752
Coeficiente ocupación de suelo	-1.336	9.201e-1	-1.452	0.148821
Viviendas	-5.553e-4	2.037e-3	-0.273	0.78554
Locales comerciales	-1.474e-2	1.653e-2	-0.892	0.373909
oficinas	-4.414e-3	7.52e-3	-0.587	0.55821
Zona censal 3_6	-2.385e-1	6.649e-1	-3.587	0.000464***
Zona censal 6_1	-8.245e-1	3.62e-1	-2.277	0.024299*
Zona censal 5_1	-1.135	3.789e-1	-2.996	0.003244**
Zona censal 7_2	-1.022	6.082e-1	-1.681	0.095065*

Fuente: Elaboración propia

Tabla 8: Primera regresión lineal múltiple resultados

Error estándar residuos	0.8436
R cuadrado múltiple	0.5896

R cuadrado ajustado	0.5331
P value	<2.2e-16

Fuente: Elaboración propia

Como se puede observar en la Tabla 7 y 8, una regresión lineal múltiple proporciona varios resultados que deben analizarse y evaluarse adecuadamente. Algunos de los valores a considerar incluyen:

En primer lugar, para analizar la regresión se debe observar el R cuadrado ajustado, que en este caso tiene en nombre Ajusted R-squared. Este es una medida de qué tan bien se ajusta el modelo de regresión lineal a los datos observados. Indica la proporción de variabilidad en la variable dependiente que es explicada por las variables independientes incluidas en el modelo (Minitab, 2019).

En este caso, el valor del R cuadrado ajustado es 0.5331, lo que significa que aproximadamente el 53.31% de la variabilidad en el precio de suelo puede ser explicada por los factores más importantes que afectan al precio de suelo incluidos en el modelo.

Un R cuadrado de 0.5331 indica que el modelo de regresión lineal múltiple utilizado explica una proporción moderada de la variabilidad en el precio de suelo. Sin embargo, esto también implica que alrededor del 46.69% de la variabilidad en el precio de suelo no puede ser explicada por las variables independientes incluidas en el modelo. (Minitab, 2019)

En segundo lugar, se debe observar el error estándar de los residuos, que en este caso este dictado por el nombre de Residual estándar error, este error es una medida de dispersión que indica la precisión con la que el modelo de regresión lineal múltiple puede predecir los valores observados de la variable dependiente.

En este caso, el valor del error estándar de los residuos es 0.8436. Esto implica que, en promedio, las predicciones del modelo de regresión lineal múltiple para el precio de suelo tienen un error estándar de 0.8436 unidades en relación con los valores reales observados (Gil, 2018).

Un error estándar de los residuos más bajo indica una mayor precisión del modelo en la predicción de los valores de la variable dependiente. Por lo tanto, un valor de 0.8436 es positivo en términos de la regresión, ya que indica que el modelo tiene una capacidad razonable para ajustarse a los datos y hacer predicciones cercanas a los valores reales (Gil, 2018).

Otro valor para considerar para evaluar la regresión son los p-value, una medida estadística que indica la probabilidad de obtener un resultado igual o más extremo que el observado, bajo la suposición de que la hipótesis nula es verdadera (Rodo, 2020)m. En el contexto de una regresión, el valor p se utiliza para evaluar la significancia estadística de una variable independiente y determinar si esta variable tiene un efecto significativo sobre la variable dependiente. En este caso, para la variable años el p-value es de 0.001362. Esto significa que la probabilidad de obtener un resultado igual o más extremo que el observado (es decir, una relación entre la variable "años" y el precio de suelo) bajo la suposición de que no hay relación real entre estas variables (hipótesis nula), es de 0.001362, o aproximadamente del 0.1362%.

En este caso particular, el p-value de 0.001362 indica que hay evidencia significativa para rechazar la hipótesis nula y afirmar que la variable "años" tiene un efecto estadísticamente significativo en el precio de suelo. Esto significa que el número de años tiene una influencia significativa en el precio de suelo y que es una variable relevante para el modelo de regresión.

Es importante tener en cuenta que el p-value no proporciona información sobre la magnitud o dirección del efecto. Solo indica si el efecto es estadísticamente significativo o no. Para obtener más información sobre la magnitud del efecto, se pueden analizar los coeficientes de regresión y otras medidas de ajuste del modelo.

El último valor para considerar son los coeficientes de las variables, estos indican en qué medida numérica afecta la variable independiente correspondiente a la variable dependiente. Como ejemplo, se utilizará la variable superficie edificada útil total que su coeficiente es de -41.94. Un coeficiente negativo significa que hay una relación inversa entre la variable independiente y la variable dependiente. En este caso, el coeficiente de -41.94 indica que, manteniendo todas las

demás variables constantes, un aumento de una unidad en la variable independiente se asocia con una disminución del 41.94% en la variable dependiente. Esta interpretación se realiza con porcentajes debido a que la variable dependiente esta transformada a logaritmo.

Es importante tener en cuenta que la interpretación del coeficiente debe considerar el contexto y las unidades de las variables involucradas. En este caso, el coeficiente de -41.94 implica que a medida que la "superficie edificada útil total" aumenta, el precio de suelo tiende a disminuir. Sin embargo, es necesario considerar otros factores y analizar el coeficiente en conjunto con el valor p.

Además, es fundamental recordar que la interpretación del coeficiente está condicionada por las características y limitaciones del conjunto de datos utilizado en la regresión. Es posible que existan otras variables relevantes que no se hayan incluido en el modelo y que podrían afectar la relación entre la "superficie edificada útil total" y el precio de suelo.

Un factor en cuenta es el estado en que se encuentra la variable tanto dependiente como independiente, esto afecta en la medida en que se interpreta el coeficiente de la variable dependiente, luego se hablará de eso cuando se obtengan los resultados de la regresión lineal múltiple final.

Una vez obtenido los resultados de la regresión con las variables obtenidas a través de paquete de caret se deben eliminar las variables que no aportan al modelo. Para esto se hicieron pruebas buscando el mejor R cuadrado ajustado posible, se eliminaron una por una hasta verificar que variables aportaban más o menos al modelo. Una vez se hizo esta prueba, las variables eliminadas fueron las siguientes: oficinas, viviendas, altura máxima y superficie edificada útil total. Estas variables no aportaban a este modelo, por lo que no se consideraron resultando la siguiente regresión lineal múltiple.

Tabla 9: Regresión lineal después de eliminar variables independientes

Coeficientes	Estimado	Error Estándar	T value	P value
--------------	----------	-------------------	---------	---------

Intercept	-1.951e2	5.751e1	-3.392	0.000898***
Cuad coef de constructibilidad	7.454e-2	2.146e-2	3.474	0.000681***
Cuad número de pisos	7.471e-4	5.16e-4	1.448	0.149865
Años	9.695e-2	2.851e-2	3.4	0.000876***
Cuad superficie edificada	6.909e-12	4.531e-11	0.152	0.879018
Densidad	2.966e-3	1.966e-3	1.509	0.133611
Estacionamientos	-1.481e-3	1.02e-3	-1.451	0.149009
Cuad superficie de terreno	-1.41e-10	6.591e-11	-2.14	0.0341*
Rasantes	4.86e-2	1.185e-2	4.1	6.91e-5***
Numero de edificaciones	8.327e-3	2.043e-2	0.408	0.684172
Coefficiente ocupación de suelo	-1.483	8.354e-1	-1.775	0.078065
Locales comerciales	-2134e-2	1.26e-2	-1.694	0.092537
Zona censal 3_6	-2.399	6.58e-1	-3.646	0.000373***
Zona censal 6_1	-8.166e-1	3.565e-1	-2.29	0.023479*
Zona censal 5_1	-1.229	3.581e-1	-3.433	0.000783***
Zona censal 7_2	-1.045	6.017e-1	-1.736	0.084695*

Fuente: Elaboración propia

Tabla 10: Regresión lineal después de eliminar variables independientes resultados

Error estándar residuos	0.8365
-------------------------	--------

R cuadrado múltiple	0.5848
R cuadrado ajustado	0.541
P value	<2.2e-16

Fuente: Elaboración propia

Como se puede observar en la Tabla 9 y 10, existe una mejora en los resultados de la regresión debido al aumento del R cuadrado ajustado y mayor significancia en los p-value resultantes de las variables independientes.

El siguiente paso es ver qué transformaciones de las variables independientes aportan mejor al modelo y perfeccionan este. Las variables que fueron analizadas por sus transformaciones fueron las siguientes: coeficiente de constructibilidad, número de pisos, año de compra de terreno, superficie edificada, superficie de terreno, rasantes y coeficiente de ocupación de suelo. Es por esto por lo que al ser 7 variables se realizaron 21 regresiones (7 variables con 3 estados) para observar que estado de cada una de las variables estudiadas era la mejor. El resultado de esta regresión final con los estados de las variables elegido son los siguientes.

Tabla 11: Regresión lineal con variables transformadas

Coefficientes	Estimado	Error Estándar	T value	P value
Intercept	-1.103e2	2.861e1	-3.856	0.000174***
Cuad coef de constructibilidad	4.819e-2	2.119e-2	2.274	0.024469*
Log número de pisos	6.834e-1	2.086e-1	3.276	0.001324**
Cuad Años	2.763e-5	7.003e-6	3.945	0.0000125***
Log superficie edificada	-2.926e-1	1.18e-1	-2.479	0.014347*
Densidad	5.38e-3	1.743e-3	3.086	0.002439**
Estacionamientos	-9.836e-4	4.076e-4	-2.413	0.01709

Superficie de terreno	-3.377e-5	9.046e-6	-3.733	0.000273***
Cuad Rasantes	3.425e-4	1.165e-4	2.94	0.003828**
Numero de edificaciones	4.76e-2	2.152e-2	2.212	0.02857*
Log Coeficiente ocupación de suelo	-6.828e-1	2.546e-1	-2.682	0.008189**
Locales comerciales	-2.055e-2	1.039e-2	-1.977	0.049935*
Zona censal 3_6	-1.981	6.398e-1	-3.096	0.002362**
Zona censal 6_1	-8795e-1	3.397e-1	-2.59	0.010611*
Zona censal 5_1	-9.942e-1	3.413e-1	-2.913	0.004157**
Zona censal 7_2	-1.044	5.71e-1	-1.828	0.084695*

Fuente: Elaboración propia

Tabla 12: Regresión lineal con variables transformadas resultados

Error estándar residuos	0.7931
R cuadrado múltiple	0.6267
R cuadrado ajustado	0.5873
P value	<2.2e-16

Fuente: Elaboración propia

En la Tabla 11 y 12 se puede observar los resultados finales de la regresión con los respectivos coeficientes por cada variable independiente. Esta regresión es altamente confiable debido a los p-value obtenidos, ya que se obtuvieron valores menores a 0.05 en todas las variables, exceptuando la zona censal 7_2, pero igualmente su valor es cercano a 0.05, por lo que se toma en consideración igualmente.

4.4 HETEROCEDASTICIDAD EN LOS DATOS

Como se mencionó anteriormente, para evaluar la evidencia de heterocedasticidad en los datos se utilizó la prueba llamada Breusch-Pagan. Los resultados fueron los siguientes:

Tabla 13: Test Breusch-Pagan

BP	80.848
df	15
p-value	4.886e-11

Fuente: Elaboración propia

En la Tabla 13 se puede observar como el p-value resultante de la prueba es igual a 4.886e-11 lo cual es significativamente menor a 0.05, concluyendo así que existe heterocedasticidad en los datos.

Existen varios métodos para tratar la heterocedasticidad de los datos, uno de ellos es el método de estimación robustos, este se utiliza para obtener estimaciones de los parámetros del modelo que son menos afectadas por valores atípicos y violaciones de los supuestos del modelo. Como se puede ver en la Ilustración 12, se presenta la regresión luego de aplicar este método.

Tabla 14: Regresión lineal método estimación robustos

Coefficientes	Estimado	Error Estándar	T value	P value
Intercept	-1.103e2	3.576e1	-3.085	0.00244**
Cuad coef de constructibilidad	4.819e-2	1.859e-2	2.592	0.01055*
Log número de pisos	6.834e-1	2.553e-1	2.677	0.00831**
Cuad Años	2.763e-5	8.753e-6	3.156	0.00195**
Log superficie edificada	-2.926e-1	1.706e-1	-1.715	0.08855.
Densidad	5.38e-3	1.803e-3	2.985	0.00334**
Estacionamientos	-9.836e-4	3.368e-4	-2.92	0.00407**

Superficie de terreno	-3.377e-5	5.933e-6	-5.692	6.94e-8***
Cuad Rasantes	3.425e-4	1.296e-4	2.643	0.00913**
Numero de edificaciones	4.76e-2	2.190e-2	2.173	0.03142*
Log Coeficiente ocupación de suelo	-6.828e-1	2.926e-1	-2.334	0.02103*
Locales comerciales	-2.055e-2	1.325e-2	-1.551	0.12304
Zona censal 3_6	-1.981	2.244	-0.883	0.37893
Zona censal 6_1	-8795e-1	6.958e-1	-1.264	0.20826
Zona censal 5_1	-9.942e-1	3.087e-1	-3.22	0.00159**
Zona censal 7_2	-1.044	8.999e-1	-1.16	0.24804

Fuente: Elaboración propia

Tabla 15: Regresión lineal método estimación robustos resultados

Error estándar residuos	0.7931
R cuadrado múltiple	0.6267
R cuadrado ajustado	0.5873
P value	<2.2e-16

Fuente: Elaboración propia

En la Tabla 14 y 15 se puede observar el resultado de la regresión lineal, luego de aplicar el método de estimación de robustos, se pueden observar cambios en los t-value como en los p-value.

4.5 ANÁLISIS DE RESULTADOS REGRESIÓN LINEAL

El resultado del R cuadrado ajustado es de 0.5873, esto indica que el modelo de regresión lineal múltiple no explica una gran proporción de la variabilidad de la variable dependiente. Esto significa que las variables independientes en el modelo no tienen un fuerte impacto en la variable dependiente y puede que existan otras variables que expliquen mejor la variabilidad observada en la variable dependiente. Es importante tener en cuenta que un valor bajo de R cuadrado ajustado

no significa necesariamente que el modelo sea inútil, sino que indica que el modelo puede ser mejorado. En este caso, se podría considerar la inclusión de otras variables independientes en el modelo o probar diferentes transformaciones de las variables existentes para mejorar el ajuste del modelo y, por lo tanto, explicar una mayor proporción de la variabilidad en la variable dependiente. Variables como la delincuencia, infraestructura y condiciones del mercado pueden ser igualmente o más influyentes en el precio del suelo que las variables estudiadas.

También existe el hecho de que se tienen pocos datos en el modelo de regresión en comparación con la cantidad de variables independientes, esto puede producir los siguientes problemas: poco poder explicativo por parte del modelo, mayor incertidumbre en los datos, overfitting y falta de representatividad del modelo. En resumen, tener pocos datos en un modelo de regresión lineal múltiple puede afectar negativamente la calidad del modelo.

A continuación, se analizarán los datos obtenidos en la regresión lineal, luego de abordar la heterocedasticidad. En primer lugar, se analizarán los resultados obtenidos en los coeficientes de las variables independientes, estos análisis varían si la variable a analizar está en logaritmo o la variable dependiente está en logaritmo. Es por esto por lo que hay que saber cómo interpretarlas.

Si la variable dependiente está en logaritmo y la variable independiente está en logaritmo, el coeficiente estimado indica el cambio porcentual en la variable dependiente (en términos de porcentaje) por un cambio porcentual en la variable independiente.

Si la variable dependiente está en logaritmo y la variable independiente no está en logaritmo, el coeficiente estimado indica el cambio porcentual en la variable dependiente por un cambio unitario en la variable independiente.

Si la variable dependiente no está en logaritmo y la variable independiente está en logaritmo, el coeficiente estimado indica el cambio porcentual en la variable dependiente por un cambio unitario en la variable independiente en términos de logaritmos.

Con esta información, los coeficientes obtenidos en la Ilustración 12 se interpretan de la siguiente manera:

- a) Zona Censal 13_6: el coeficiente estimado de -1.981 indica que, manteniendo las demás variables constantes, los terrenos que pertenezcan a esta zona censal tienen un 198.1% menos en su precio del suelo en comparación con los terrenos que no pertenezcan a la zona. Esta zona se encuentra en la calle Presidente Errazuriz y alberga varios colegios y universidades. La presencia de instituciones educativas y la ubicación pueden influir en la percepción del valor del suelo en esta área, lo que se refleja en la reducción del precio.
- b) Zona Censal 7_2: el coeficiente estimado es igual a -1.044 indica que, manteniendo las demás variables constantes, los terrenos que pertenezcan a esta zona censal tienen un 104.4% menos en su precio del suelo en comparación con los terrenos que no pertenezcan a la zona. Esta zona pertenece a Apoquindo específicamente por el centro comercial Apumanque, en ella se pueden observar una gran cantidad de locales comerciales y hospitales por la zona.
- c) Zona Censal 15_1: el coeficiente estimado es igual a -0.9942 indica que, manteniendo las demás variables constantes, los terrenos que pertenezcan a esta zona censal tienen un 99.42% menos en su precio del suelo en comparación con los terrenos que no pertenezcan a la zona. Esta zona pertenece a San Carlos de Apoquindo, en ella se encuentra una diversa cantidad de locales comerciales, colegios y universidades.
- d) Zona Censal 16_1: el coeficiente estimado es igual a -0.8795 indica que, manteniendo las demás variables constantes, los terrenos que pertenezcan a esta zona censal tienen un 87.95% menos en su precio del suelo en comparación con los terrenos que no pertenezcan a la zona. Esta zona censal pertenece al barrio los Dominicos, en ella se encuentra el metro de los Dominicos.
- e) Número de pisos: el coeficiente estimado es igual a 0.6834, significa que un aumento del 1% en la variable independiente, manteniendo las demás variables constantes, se asocia con un aumento del 68.34% en la variable dependiente. En zonas donde la densidad de construcción es un factor importante, como áreas urbanas densamente pobladas, un mayor número de pisos puede permitir una mayor rentabilidad de la inversión, lo que se refleja en un precio del suelo más alto.

- f) Coeficiente ocupación de suelo: el coeficiente estimado es igual a -0.6828 , significa que un aumento del 1% en la variable independiente, manteniendo las demás variables constantes, se asocia con una disminución del 68.28% en la variable dependiente. Este resultado enfatiza la importancia del uso del suelo en la valorización. Un coeficiente de ocupación de suelo más alto implica que se puede construir más en el terreno, lo que generalmente se asocia con propiedades más densas o edificios más altos. Sin embargo, esto también puede indicar una mayor oferta de terrenos con características similares, lo que puede ejercer presión a la baja sobre los precios.
- g) Superficie edificada: el coeficiente estimado es igual a -0.2926 , significa que un aumento del 1% en la variable independiente, manteniendo las demás variables constantes, se asocia con una disminución del 29.26% en la variable dependiente. Una posible explicación de esta relación inversa es que los proyectos con superficies edificadas más grandes pueden estar asociados con mayores costos de construcción y desarrollo. Esto podría disminuir la rentabilidad general del proyecto y, en consecuencia, reducir el precio del suelo.
- h) Coeficiente de constructibilidad: el coeficiente estimado es igual a 0.04819 , significa que un aumento de una unidad en la variable independiente, manteniendo las demás variables constantes, se asocia con un aumento del 4.819% en la variable dependiente. A pesar de que existe un gran aumento en la variable dependiente, hay que tener en cuenta que el coeficiente de constructibilidad es un dato que no varía mucho en términos de unidades. La desviación estándar de la variable independiente⁴, se muestra un valor de 1.233, lo que se explica como que en ese valor varía la variable a lo largo de los datos.
- i) Número de edificaciones: el coeficiente estimado es igual a 0.0476 , significa que un aumento de una unidad en la variable independiente, manteniendo las demás variables constantes, se asocia con un aumento de 4.76% en la variable dependiente. Se puede interpretar como que las ubicaciones con mayor cantidad de edificaciones ya construidas pueden tener un mayor potencial de desarrollo inmobiliario. Esto podría atraer a inversores que buscan oportunidades para expandir o mejorar las edificaciones existentes, lo que aumentaría la demanda y, por ende, el precio del suelo.
- j) Locales comerciales: el coeficiente estimado es igual a -0.02055 , significa que un aumento de una unidad en la variable independiente, manteniendo las demás variables constantes,

⁴ Desviación estándar en Anexo 10

se asocia con una disminución de 2.055 % en la variable dependiente. Este resultado sugiere que, en la ubicación analizada, la inclusión de locales comerciales en una edificación puede tener un efecto ligeramente negativo en el precio del suelo. Esto podría deberse a que los terrenos destinados a fines comerciales pueden tener un valor unitario del suelo menor en comparación con terrenos destinados a otros usos, como viviendas.

- k) Densidad: el coeficiente estimado es igual a 0.00538, significa que un aumento de una unidad en la variable independiente se asocia con un aumento del 0.538% en la variable dependiente. Este valor se interpreta como que igualmente tiene alta significancia en el valor del suelo, esto se debe a que la variable densidad tiene una alta desviación estándar de 95.25⁵ por lo que la diferencia de valores es grande.
- l) Estacionamientos: el coeficiente estimado es igual a -0.0009836, significa que un aumento de una unidad en la variable independiente, manteniendo las demás variables constantes, se asocia con una disminución de 0.09836 % en la variable dependiente. Este resultado indica que, en la ubicación analizada, una mayor cantidad de estacionamientos en una edificación se asocia con una leve disminución en el precio del suelo. Esto podría deberse a que, en ciertos contextos, los terrenos destinados a estacionamientos pueden tener un valor unitario del suelo menor en comparación con terrenos destinados a otros usos más intensivos.
- m) Rasantes: el coeficiente estimado es igual a 0.0003425, significa que un aumento de una unidad en la variable independiente, manteniendo las demás variables constantes, se asocia con un aumento de 0.03425 % en la variable dependiente. Esta variable no es muy significativa, ya que, por lo general, en la vista de datos, se obtenían muchos valores iguales y cercanos al 70° en el ángulo de rasantes.
- n) Superficie de terreno: el coeficiente estimado es igual a -0.00003377, significa que un aumento de una unidad en la variable independiente, manteniendo las demás variables constantes, se asocia con una disminución de 0.003377 % en la variable dependiente. Esta variable no tiene un impacto muy significativo en el precio del suelo en el contexto analizado. Esto podría indicar que, en este conjunto de datos, los cambios en el tamaño del terreno tienen un efecto mínimo en la valorización de los terrenos.

⁵ Desviación estándar en Anexo 11

- o) Años de compra de terreno: el coeficiente estimado es igual a 0.00002763, significa que un aumento de una unidad en la variable independiente se asocia con un aumento del 0.002763% en la variable dependiente. Es importante destacar que, al tener la variable al cuadrado, las relaciones cuadráticas pueden indicar que, inicialmente, un aumento en los años de compra de terreno puede tener un efecto limitado en el precio del suelo, pero a medida que los años aumentan aún más, ese efecto puede volverse más pronunciado.

El análisis de las variables en el contexto de su uso como herramientas para una inmobiliaria en la valoración de terrenos proporciona información valiosa para la toma de decisiones de inversión. Estas variables, en conjunto, ofrecen una visión integral de cómo diversos factores influyen en el precio del suelo, lo que es esencial para entender y prever tendencias en el mercado inmobiliario.

La ubicación, como uno de los factores principales, puede indicar áreas geográficas más propicias para la inversión. Además, se observa cómo características de las edificaciones, como la altura y el uso del suelo, pueden afectar la valorización. Por ejemplo, un mayor número de pisos puede aumentar el precio del suelo, mientras que un mayor coeficiente de ocupación de suelo podría tener el efecto contrario. La presencia de locales comerciales y estacionamientos también juega un papel en la valoración, y los coeficientes estimados revelan cómo estos factores pueden influir en el precio del suelo.

En este estudio, se compararon los valores reales de los precios de suelos con las predicciones generadas por el modelo de regresión lineal. Se seleccionaron aleatoriamente 14 conjuntos de datos para esta comparación⁶. Esto se hizo para evaluar cuán precisamente el modelo puede predecir los precios reales del suelo y determinar si es una herramienta confiable para la toma de decisiones en bienes raíces. Los resultados son los siguientes:

- A) El promedio de los resultados de la regresión lineal es de 3.1378.
- B) El promedio de los valores reales de suelo es de 3.2486.
- C) El promedio de la diferencia absoluta entre los valores reales y los valores predichos por la regresión lineal es de 0.3153.

⁶ Resultados en Anexo 12

La diferencia absoluta promedio de 0.3153 indica que, en promedio, las predicciones de la regresión lineal se desvían en aproximadamente 0.3153 unidades de los valores reales de suelo. Si el propósito es tomar decisiones de inversión o desarrollo inmobiliario, entonces una diferencia de 0.3153 podría ser significativa, ya que podría afectar la rentabilidad de un proyecto. Sin embargo, si el análisis es más descriptivo o de investigación, esta diferencia podría ser aceptable.

4.6 ANÁLISIS DE AUTOCORRELACIÓN Y MULTICOLINEALIDAD

En el estudio realizado, se utilizó la prueba de Durbin-Watson para evaluar la presencia de autocorrelación en los residuos de un modelo de regresión lineal. El objetivo de esta prueba fue examinar si existía alguna relación sistemática entre los residuos del modelo y la variable dependiente, que en este caso es el precio del suelo.

La prueba de Durbin-Watson se basa en la suposición de que los residuos son independientes entre sí. Si hay autocorrelación en los residuos, significa que existe alguna dependencia serial entre ellos, lo que puede afectar la precisión y validez del modelo de regresión.

En RStudio, se aplicó la función "durbinWatsonTest()" para realizar la prueba de Durbin-Watson. Esta función calcula la estadística de prueba DW y el valor p asociado. Como se puede observar en la tabla 16 el resultado obtenido fue un valor de DW igual a 1.9593 y un valor p de 0.373.

Tabla 16: Prueba Durbin-Watson

Valor DW	1.9593
Valor p	0.373

Fuente: Elaboración propia

La interpretación de los resultados se realizó de la siguiente manera: un valor de DW cercano a 2 indica la ausencia de autocorrelación significativa en los residuos. En este caso, el valor de DW obtenido sugiere que los residuos del modelo de regresión lineal del precio del suelo no presentan

una dependencia serial importante. Además, el valor p de 0.373 indica que no hay suficiente evidencia para rechazar la hipótesis nula de no autocorrelación.

Estos hallazgos sugieren que el modelo de regresión lineal utilizado en el estudio es apropiado y que los residuos no muestran patrones sistemáticos de autocorrelación. Por lo tanto, se puede asumir que los supuestos de independencia de los residuos se cumplen en el análisis del precio del suelo.

Para analizar la multicolinealidad se realizó una prueba de factor de inflación de la varianza (VIF, por sus siglas en inglés) en RStudio para evaluar la presencia de colinealidad en un modelo de regresión lineal. El VIF es una medida que cuantifica la relación entre una variable independiente específica y las demás variables independientes en el modelo. Indica cuánto se incrementa la varianza de un coeficiente de regresión debido a la correlación con otras variables independientes. Un valor de VIF igual a 1 indica ausencia completa de colinealidad, mientras que valores mayores a 1 indican la presencia de colinealidad, donde valores más altos indican una mayor intensidad de colinealidad.

Los resultados del VIF fueron:

1. Zona censal7_2: 1.000143
2. Zona censal 13_6: 1.000126
3. Zona censal 16_1: 1.001438
4. Zona censal 15_1: 1.260241
5. Locales comerciales: 1.727747
6. Años: 2.090051
7. Superficie de terreno: 2.457268
8. Densidad: 4.759526
9. Estacionamientos: 7.155338
10. Número de edificaciones: 10.956951
11. Coeficiente ocupación de suelo: 11.176089
12. Rasantes: 13.374192
13. Coeficiente de constructibilidad: 14.899038

14. Superficie edificada: 24.864391

15. Número de pisos: 34.983028

Las variables Zona censal7_2, Zona censal 13_6, Zona censal 15_1, Locales comerciales y Zona censal 16_1 tienen valores de VIF muy cercanos a 1, lo cual sugiere que no hay colinealidad entre estas variables y las demás.

En general, se considera que un VIF menor a 5 indica una baja presencia de colinealidad, mientras que valores superiores a 5 pueden indicar una alta colinealidad. En este caso, las variables: años, superficie de terreno y densidad tienen valores de VIF mayores a 2, pero menores a 5, lo cual sugiere una presencia moderada de colinealidad.

Sin embargo, las variables número de edificaciones, coeficiente ocupación de suelo, rasantes, coeficiente de constructibilidad, superficie edificada y número de pisos tienen valores de VIF bastante elevados, todos por encima de 10. Esto indica una alta colinealidad entre estas variables y las demás en el modelo de regresión.

En muchos casos, la colinealidad puede surgir debido a la falta de información sobre ciertas variables que son relevantes para el modelo. En lugar de eliminar o transformar variables existentes, una estrategia efectiva es recopilar datos adicionales que puedan ayudar a resolver el problema de la colinealidad. Algunas consideraciones clave para recopilar estos datos son: identificar las variables faltantes, obtener datos confiables, analizar el impacto que tienen y validar el modelo.

5. CONCLUSIONES Y RECOMENDACIONES

5.1 CONCLUSIONES GENERALES

Los hallazgos de este trabajo resaltan la influencia significativa de diversas variables en el precio del suelo. Las zonas censales, la densidad de construcción, el coeficiente de ocupación de suelo, el número de pisos y la presencia de locales comerciales han demostrado ser factores clave en la valoración del suelo. Estos resultados proporcionan una visión detallada de cómo las características del vecindario y las variables de construcción influyen en el valor del suelo.

Para las inmobiliarias, esta información es esencial. Les permite tomar decisiones financieras sólidas y estratégicas que afectan directamente la rentabilidad y el éxito de sus proyectos. La capacidad de ajustar estrategias de precios, diversificar su portafolio, obtener respaldo financiero y evaluar los riesgos asociados con un proyecto se ve significativamente mejorada gracias a este modelo.

La comprensión de cómo estas variables afectan el precio del suelo no solo mejora la rentabilidad de los proyectos inmobiliarios, sino que también contribuye a una toma de decisiones más informada y precisa en las carteras de inversión. Esto es crucial en un mercado inmobiliario en constante cambio y competitivo.

Con base en estas conclusiones, se puede afirmar que el modelo desarrollado en la memoria permitió comprender los factores que influyen en la determinación del precio del suelo y medir correctamente los incrementos en los precios controlados por sus características, a pesar de que el R cuadrado ajustado de nuestro modelo es bajo.

El enfoque econométrico utilizado proporcionó una metodología sólida para analizar y explicar el comportamiento del mercado inmobiliario y su relación con el precio del suelo.

Además, es importante destacar que, si bien este modelo ha demostrado ser valioso para comprender una serie de factores que influyen en el precio del suelo, el coeficiente de determinación ajustado (R^2 ajustado) indica que aún existe una proporción significativa de la variabilidad en el precio del suelo que no se ha explicado completamente. Esto abre la puerta a futuras investigaciones y análisis más detallados sobre otras variables que pueden influir en el precio del suelo y que no se han considerado en este estudio. Estas investigaciones adicionales pueden proporcionar una comprensión aún más profunda y precisa del mercado inmobiliario, lo que beneficiará aún más a las empresas inmobiliarias en sus decisiones estratégicas.

5.2 CONCLUSIONES ESPECÍFICAS

La infraestructura de un terreno es un factor crítico por considerar para las inmobiliarias al utilizar esta información. Una infraestructura sólida y de alta calidad, que incluye suministros básicos como agua potable, alcantarillado, y buenas vías de comunicación, puede aumentar significativamente el valor del suelo en una ubicación. Esto se debe a que una infraestructura bien desarrollada mejora la calidad de vida de los residentes y hace que la zona sea más atractiva para vivir. Por otro lado, si la infraestructura es inadecuada o está en mal estado, esto puede tener un impacto negativo en el precio del suelo. Los terrenos ubicados en áreas con servicios de mala calidad o infraestructura deficiente suelen tener un valor más bajo, lo que puede afectar la rentabilidad de los proyectos inmobiliarios. Por lo tanto, las inmobiliarias deben considerar cuidadosamente la calidad de la infraestructura al tomar decisiones estratégicas sobre dónde invertir y desarrollar proyectos, ya que esto puede tener un impacto significativo en la rentabilidad a largo plazo.

La delincuencia es una variable que puede depender del suelo, ya que el entorno en el que se encuentra un terreno puede influir en la tasa de delincuencia de la zona. Para las inmobiliarias, esto reviste especial importancia, ya que la seguridad es un criterio fundamental para la elección de un lugar para vivir o establecer negocios. Una alta tasa de delincuencia puede reducir significativamente el valor del suelo, ya que muchas personas pueden percibir la zona como insegura y poco atractiva. Por otro lado, en áreas con bajos índices de delincuencia, el valor del

suelo tiende a aumentar, ya que estas zonas son más deseables tanto para residentes como para inversores.

Es importante tener muchos datos en una regresión lineal múltiple porque una muestra grande de datos proporciona una mejor estimación de la relación entre las variables independientes y la variable dependiente. Cuando se tiene una gran cantidad de datos, se pueden detectar relaciones más precisas y complejas entre las variables, lo que puede mejorar la precisión de la regresión lineal múltiple.

Utilizando este modelo, se pueden realizar pronósticos y estimaciones de los precios del suelo con base en las características específicas de un terreno dado. Al considerar las variables estudiadas, se pueden obtener predicciones más precisas y realistas de los precios para estudios realizados por inmobiliarias. También pueden ser útiles para la elaboración de políticas públicas relacionadas con el uso del suelo y la planificación urbana.

Es importante tener en cuenta que las predicciones basadas en el modelo econométrico no son infalibles y están sujetas a ciertas limitaciones y supuestos del modelo. Sin embargo, al utilizar un enfoque riguroso y una metodología sólida, se pueden obtener estimaciones más confiables y fundamentadas en comparación con simples intuiciones o estimaciones subjetivas.

En resumen, se logró comprender cómo influyen en el precio del suelo. El modelo desarrollado tomó en cuenta factores como la ubicación, número de pisos, coeficiente de ocupación, superficie edificada, entre otros. A pesar de que el modelo no es perfecto, proporciona información valiosa para tomar decisiones informadas en el mercado inmobiliario. Las predicciones basadas en el modelo pueden ser utilizadas por inversionistas, desarrolladores y planificadores urbanos, y también pueden contribuir a la elaboración de políticas públicas relacionadas con el uso del suelo y la planificación urbana.

5.3 RECOMENDACIONES FUTURAS

Las recomendaciones de investigación se centran en proporcionar orientación para las decisiones que tomarán las empresas relacionadas con la industria del precio del suelo. También se busca mejorar el modelo de regresión lineal existente mediante la inclusión de más variables con el objetivo de hacer que el modelo se ajuste mejor a la realidad. Además, se sugiere incorporar más datos de precios de suelo para evitar problemas de dimensionalidad en el análisis.

En primer lugar, se recomienda aumentar la base de datos, esto para así mejorar la regresión lineal, ya que el R cuadrado aumentaría, produciendo así una mayor significancia en los datos para los resultados obtenidos. Las variables para considerar son las siguientes:

- a) Tasa de interés hipotecaria: La tasa de interés en los préstamos hipotecarios puede influir en la demanda de propiedades. Un aumento en las tasas de interés puede reducir la capacidad de las personas para comprar viviendas, lo que podría afectar los precios del suelo.
- b) Distribución de zonas verdes: La cantidad de áreas verdes o parques en una zona puede afectar la calidad de vida y, por lo tanto, los precios del suelo.
- c) Índices de criminalidad: La seguridad es un factor importante para los compradores de viviendas. Las tasas de criminalidad en una zona pueden afectar la percepción de seguridad y, en última instancia, los precios del suelo.
- d) Crecimiento económico futuro: Factores como los planes de desarrollo urbano, la construcción de nuevas infraestructuras o la llegada de empresas pueden indicar un aumento en la demanda de viviendas y, por lo tanto, en los precios del suelo.

A estas variables se les debe realizar un análisis cuidadoso para determinar su relevancia y su efecto potencial en el precio del suelo. También se deben realizar pruebas de multicolinealidad y ajustes al modelo a medida que agreguen más variables para asegurar de que el modelo sea adecuado y preciso.

En segundo lugar, se recomienda aumentar el tamaño de la observación. Una base de datos más grande puede tener un impacto significativo en la precisión y robustez de un modelo de regresión

lineal. Si la base de datos hubiera tenido más observaciones, habrías tenido una visión más completa y precisa del mercado inmobiliario en tu área de interés. Esto habría permitido:

- a) Identificar patrones y tendencias con mayor confianza.
- b) Establecer relaciones estadísticas más sólidas entre las variables independientes y el precio del suelo.
- c) Refinar y validar el modelo de regresión lineal con una mayor cantidad de datos de prueba.
- d) Realizar análisis más detallados, como la evaluación de precios en subregiones específicas.

Por lo tanto, considerando todo lo anterior, una recomendación clave sería explorar el uso del web scraping para recolectar datos de manera automatizada y así obtener una base de datos más amplia y rica en observaciones. Esta estrategia habría permitido capturar información de diversas fuentes en línea, incluyendo diferentes ubicaciones geográficas y períodos de tiempo, lo que habría enriquecido significativamente el conjunto de datos.

6. BIBLIOGRAFÍA

Amat, J., 2018. *Machine Learning con R y caret*, : ciencia de datos.

Asturias Corporación Universitaria, s.f. *Multicolinealidad, Heterocedasticidad, utocorrelación*, :

Berrendero, J., 2016. *Regresión lineal simple*, : EStudio.

Carmona, J., s.f. *desviacion estandar y tipica*. ,

CFI , 2022. *Variance Inflation Factor (VIF)*, :

Ciencia sin Sesos, 2014. *Engañando a Gauss*. ,

Cifuentes, A., 2020. *DETERMINANTES DE LOS PRECIOS EN EL MERCADO INMOBILIARIO SOBRE LA BASE DE UN ÍNDICE DE CUALIDADES DE LA VIVIENDA*, :

DÍAZ, G. C., 2014. *PRECIO DEL SUELO Y METODOLOGÍAS DE AVALUACIÓN*., Santiago: Universidad de Chile.

Engel & Völkers Blog, 2015. *Engel & Völkers*. [En línea]
Available at: <https://www.engelvoelkers.com/es/blog/perspectivas-inmobiliarias/tendencias-del-mercado/inmobiliaria-en-chile-un-mercado-muy-demandado/>

Esparza, E., 2016. *pisos*. [En línea]
Available at: <https://www.pisos.com/aldia/como-afecta-el-precio-del-suelo-en-el-precio-final-de-la-vivienda/71755/>

Espinoza, E., 2018. *Las variables y su operacionalización en la investigación educativa*.. ,

Espinoza, M., 2017. *ANÁLISIS DE CORRELACION Y REGRESION LINEAL EN FUNCION DE ESTUDIANTES MATRICULADOS DE LA UNIVERSIDAD ECOTEC*, : Congreso Científico Internacional Sociedad del Conocimiento: Retos y Perspectivas.

Figuroa, E., 2006. *Determinantes del precio de la vivienda en Santiago: Una estimación hedónica*, :

finmodelslab, 2023. *Impulsar ganancias del desarrollo de la tierra*, :
Flores, E., 2016. *PRUEBA DE BREUSCH PAGAN-GODFREY, TEST DE HETEROCEDASTICIDAD*, : Prezi.

Gil, C., 2018. *REGRESIÓN LINEAL SIMPLE*, :
Guerrero, A., 2016. *Una evaluación del algoritmo LVQ en una colección de texto*, :
Revista Cubana de Ciencias Informáticas.

Guzman, L. A., 2020. *Land Use Policy*, :

Jara, M. M., 2008. *HETEROSCEDASTICIDAD*, : La Evolución De La Econometria.
Kuhn, M., 2019. *The caret package*, :

M, G., 2021. *El mostrador*. [En línea]
Available at: <https://www.elmostrador.cl/generacion-m/2021/03/03/precio-de-suelo-para-proyectos-de-densificacion-en-el-gran-santiago-aumenta-un-218-en-la-ultima-decada/#:~:text=En%20ese%20contexto%2C%20el%20reporte,crecimiento%20promedio%20de%2015%25%20anual>
[Último acceso: 7 Septiembre 2022].

Minitab, 2019. *Análisis de Regresión*, :
Municipalidad de Las Condes, 2022. *CUENTA GESTIÓN MUNICIPAL 2021 - 2022*, :

Novalés, A., 2010. *Análisis de Regresión*, :

Parrado, E., 2005. *Evolución de los precios de vivienda*, :

Ramírez, V., s.f. *AUTOCORRELACIÓN*, : Material econometria.

Ramón, G., s.f. *Correlación entre variables*, :

Rodo, P., 2020. *Valor-p*, :

Sagner, A., 2011. *DETERMINANTES DEL PRECIO DE VIVIENDAS EN LA REGIÓN METROPOLITANA DE CHILE*, :

Serey, D., 2019. *El Mostrador*. [En línea]
Available at: <https://www.elmostrador.cl/generacion-m/2021/03/03/precio-de-suelo-para-proyectos-de-densificacion-en-el-gran-santiago-aumenta-un-218-en-la-ultima-decada/#:~:text=En%20ese%20contexto%2C%20el%20reporte,crecimiento%20promedio%20de%2015%25%20anual>

Sii, 2022. *¿Qué significan los dígitos que conforman el Rol de una propiedad?*, :
Socovesa, 2023. *¿Qué es valor de suelo y cómo afecta los proyectos inmobiliarios?*, :

Torrealba, G. E., 2019. *PRINCIPIOS DE VIABILIDAD DE PROYECTOS INMOBILIARIOS*, :

Vela, F., 2010. *Normalidad de los errores*. ,

7. ANEXOS

ANEXO 1. PERMISOS DE EDIFICACIÓN LEVANTADOS

TIPO DE NORMA	DENOMINACIÓN	NÚMERO	FECHA	INDICACIÓN DEL MEDIO Y FORMA DE PUBLICIDAD (SEGÚN ART.45 Y SIGUIENTES LEY 19.880)	FECHA DE PUBLICACIÓN	TIENE EFECTOS GENERALES	FECHA ÚLTIMA ACTUALIZACIÓN (DD/MM/AAAA), SI CORRESPONDE A ACTOS Y RESOLUCIONES CON EFECTOS GENERALES	BREVE DESCRIPCIÓN	ENLACE A LA PUBLICACIÓN O ARCHIVO CORRESPONDIENTE	PERMISO UTILIZABLE
1	PERMISO DE EDIFICACION	176	22-12-2021	SITIO WEB DEL ORGANISMO	12-01-2021	NO	SIN MODIFICACIÓN	CONDOMINIO TIPO A DE 4 VIVIENDAS EN CAMINO LA FUENTE 13978		SI
2	PERMISO DE EDIFICACION	169	07-12-2021	SITIO WEB DEL ORGANISMO	12-01-2021	NO	SIN MODIFICACIÓN	PARROQUIA NUESTRA SEÑORA DE LOS ANGELES EN EL GOLF 141 (EX-		SI
3	PERMISO DE EDIFICACION	168	07-12-2021	SITIO WEB DEL ORGANISMO	12-01-2021	NO	SIN MODIFICACIÓN	MULTICANCHA TECHADA COLEGIO ACOGIDO AL ART. 124 DE LA LEY		SI
4	PERMISO DE EDIFICACION	167	06-12-2021	SITIO WEB DEL ORGANISMO	12-01-2021	NO	SIN MODIFICACIÓN	CONJUNTO DE 19 VIVIENDAS UNFAMILIAR DE 2 PISOS EN LA		SI
5	PERMISO DE EDIFICACION	162	23-11-2021	SITIO WEB DEL ORGANISMO	10-12-2021	NO	SIN MODIFICACIÓN	CONDOMINIO DE 23 VIVIENDAS UNFAMILIARES EN LA QUEBARADA		SI
6	PERMISO DE EDIFICACION	161	23-11-2021	SITIO WEB DEL ORGANISMO	10-12-2021	NO	SIN MODIFICACIÓN	CONDOMINIO TIPO A DE 26 VIVIENDAS EN EL ALTO 794.		SI
7	PERMISO DE EDIFICACION	159	19-11-2021	SITIO WEB DEL ORGANISMO	10-12-2021	NO	SIN MODIFICACIÓN	CONDOMINIO TIPO A DE 2 VIVIENDAS EN PADRE ERRAZURIZ 7903.		SI
8	PERMISO DE EDIFICACION	157	17-11-2021	SITIO WEB DEL ORGANISMO	10-12-2021	NO	SIN MODIFICACIÓN	UN EDIFICIO 7 PISOS + 5 SUBTERRANEOS EN LAS CONDES		SI
9	PERMISO DE EDIFICACION	155	12-11-2021	SITIO WEB DEL ORGANISMO	10-12-2021	NO	SIN MODIFICACIÓN	CINCO VIVIENDAS DE 2 PISOS + SUBTERRANEO EN CONDOMINIO TIPO		SI
10	PERMISO DE EDIFICACION	153	10-11-2021	SITIO WEB DEL ORGANISMO	10-12-2021	NO	SIN MODIFICACIÓN	CONJUNTO ARMÓNICO DE DOS EDIFICIOS, EN APOQUINDO 6000 /		SI
11	PERMISO DE EDIFICACION	149	29-10-2021	SITIO WEB DEL ORGANISMO	11-11-2021	NO	SIN MODIFICACIÓN	CONDOMINIO TIPO A DE 5 VIVIENDAS DE 3 PISOS Y 1 SUBTERRANEO EN		SI
12	PERMISO DE EDIFICACION	145	27-10-2021	SITIO WEB DEL ORGANISMO	11-11-2021	NO	SIN MODIFICACIÓN	CONDOMINIO DE 4 VIVIENDAS DE 3 PISOS Y 1 SUBTERRANEO EN		SI
13	PERMISO DE EDIFICACION	144	27-10-2021	SITIO WEB DEL ORGANISMO	11-11-2021	NO	SIN MODIFICACIÓN	CONDOMINIO TIPO A DE 5 VIVIENDAS DE 3 PISOS Y 1 SUBTERRANEO EN		SI
14	PERMISO DE EDIFICACION	139	13-10-2021	SITIO WEB DEL ORGANISMO	11-11-2021	NO	SIN MODIFICACIÓN	CONDOMINIO TIPO A DE CUATRO VIVIENDAS DE DOS PISOS Y UN		SI
15	PERMISO DE EDIFICACION	132	24/09/21	SITIO WEB DEL ORGANISMO	13-10-2021	NO	SIN MODIFICACIÓN	CONJUNTO ARMÓNICO CONFORMADO POR 2 EDIFICIOS 2		SI
16	PERMISO DE EDIFICACION	120	08-09-2021	SITIO WEB DEL ORGANISMO	13-10-2021	NO	SIN MODIFICACIÓN	16 VIVIENDAS 2 PISOS Y UN SUBTERRANEO EN CAMINO FRAY		SI
17	PERMISO DE EDIFICACION	110	16-08-2021	SITIO WEB DEL ORGANISMO	09-09-2021	NO	SIN MODIFICACIÓN	2 VIVIENDAS NUEVAS EN POLONIA 360.		SI

ANEXO 2. DOCUMENTOS DE TRANSACCIÓN

cod mz	cod pr	fojas	num	año	comprador	rut comp	vendedor	rut vend	fecha inscripcio	precio UF	precio \$	fecha escritur
2	2640	49	4386	64097	2021	INMOBILIARIA LA FUENTE LIMITADA	77.303.118-5	RUIZ LUCIRAIL	15-06-2021	36500		27-04-2021
3	1204	197	51615	75373	2021	INMOBILIARIA LOMAS DE ASIS LIMITADA	86.763.950-8	LA FUNDACION CULTURAL DE LAS CONDES	15-07-2021	114514		29-06-2021
4	1204	190	51614	75372	2021	INMOBILIARIA LOMAS DE ASIS LIMITADA	86.763.950-8	LA FUNDACION CULTURAL DE LAS CONDES	15-07-2021	138379		29-06-2021
5	1204	201	52536	76740	2021	INMOBILIARIA LOMAS DE ASIS LIMITADA	86.763.950-8	LA FUNDACION CULTURAL DE LAS CONDES	20-07-2021	156430		29-06-2021
6	2164	6	3060	19645	2021	REINENHAUS SPA	77.287.304-0	TELIX ERGAS RITA CLARA, HENRICH GOLDSCHMIDT EDGAR ELIAS	29-02-2021		\$363.695.125	21-01-2021
7	1597	10	006671	7696	2002	LA ORACION SPA	77.133.239-7	RENTAS INMOBILIARIAS SOLARO SPA	18-11-2020	10750		10-11-2020
8	1597	11	17800	18850	1982	LA ORACION SPA	77.133.239-7	TESTA BAJAR REGINA	18-11-2020	70875		09-11-2020
9	1597	12	70427	67437	2005	LA ORACION SPA	77.133.239-7	INMOBILIARIA LOS TRIGALES LIMITADA	19-11-2020	6455		09-11-2020
10	1597	21	58366	78457	2020	LA ORACION SPA	77.133.239-7	RENTAS INMOBILIARIAS SOLARO SPA	18-11-2020	10500		10-11-2020
11	1597	22	64372	46344	2000	LA ORACION SPA	77.133.239-7	DIAZ PALMA LEONOR DEL CARMEN	18-11-2020	10172		09-11-2020
12	1597	23	49350	47070	2003	LA ORACION SPA	77.133.239-7	PALMA DEL RIO MARIA CECILIA	18-11-2020	8346		09-11-2020
13	1597	24	15692	24819	2008	LA ORACION SPA	77.133.239-7	MEDINA PARRA CARMEN PAZ	18-11-2020	12000		09-11-2020
14	1597	26	48793	70963	2019	LA ORACION SPA	77.133.239-7	MEDINA ZANETTA SERGIO JULIO, MEDINA PARRA SERGIO RODRIGO, MEDINA PARRA LORETO ELENA, MEDINA PARRA CARMEN PAZ	18-11-2020	12200		09-11-2020
15	1597	27	090089	51371	2000	LA ORACION SPA	77.133.239-7	RENTAS INMOBILIARIAS SOLARO SPA	18-11-2020	10870		10-11-2020
16	1597	28	28841	32384	1986	LA ORACION SPA	77.133.239-7	RENTAS INMOBILIARIAS SOLARO SPA	18-11-2020	11270		10-11-2020
17	1276	25	44260	64733	2021	CONSTRUCTORA E INMOBILIARIA TOWNHOUSE ISABEL SPA	77.346.347-6	VIVAMERES SALDIAS MARIA DEL CARMEN, VIVAMERES SALDIAS LUIS MAURICIO, VIVAMERES SALDIAS JERARDO, VIVAMERES SALDIAS NIEVES	16-06-2021	25200		20-05-2021
18	729	7	10610	14051	1970	GESTION INMOBILIARIA ORIENTE SPA		DESARROLLOS INMOBILIARIOS S.A.	16-01-2019		\$20.123.066.792	10-01-2019
19	1144	12	47002	68715	2021	INVERSIONES EL ALBA SPA	77.337.447-3	LAIDES EL ALBA SPA	29-06-2021		\$812.360.790	26-05-2021
20	274	9	26335	16371	1995	INMOBILIARIA POLONIA SPA	77.294.629-5	PORTALES ZEGERES LUC MARIA	02-07-2021	40685		16-06-2021
21	554	7	5833	85609	2006	INVERSIONES DEL INCA SPA	77.337.446-5	RETAMAL ALMAGA JOSE LUIS	07-06-2021	26000		28-04-2021
22	2690	37	76546	11074	2021	MUÑOZ MARRIN BENJAMIN ANTONIO	17.089.752-1	INVERSIONES LA PINTA LIMITADA	05-11-2021	40814,85		31-08-2021
23	508	1	59334	55661	2002	INMOBILIARIA Y RENTAS APOQUINDO SAN PASCUAL SPA	77.139.351-9	EUROAMERICA INVERSIONES E INMOBILIARIA S.A.	30-03-2021	302224.3541		12-03-2021
24	508	4	22642	33318	2021	INMOBILIARIA Y RENTAS APOQUINDO SAN PASCUAL SPA	77.139.351-9	EUROAMERICA INVERSIONES E INMOBILIARIA S.A.	26-03-2021	69040		12-03-2021
25	508	5	51621	41212	1983	INMOBILIARIA Y RENTAS APOQUINDO SAN PASCUAL SPA	77.139.351-9	EUROAMERICA INVERSIONES E INMOBILIARIA S.A.	30-03-2021	68540		12-03-2021
26	508	10	15669	18439	1983	INMOBILIARIA Y RENTAS APOQUINDO SAN PASCUAL SPA	77.139.351-9	AGRICOLA LA ARBOLEDA LIMITADA	01-04-2021	140892.89223		12-03-2021
27	508	11	48801	79569	2006	INMOBILIARIA Y RENTAS APOQUINDO SAN PASCUAL SPA	77.139.351-9	MARIA DE LA LUZ ANGELICA OYARZUN MARIN	10-02-2021	53000		18-01-2021
28	508	12	58886	63470	2006	INMOBILIARIA Y RENTAS APOQUINDO SAN PASCUAL SPA	77.139.351-9	RAMON VALENTE Y COMPAÑIA LIMITADA	12-03-2021	43000		18-01-2021
29	2430	163	11740	17869	2012	INMOBILIARIA FRAY MARTIN LIMITADA		INMOBILIARIA ESCORIAL LIMITADA	19-04-2018			31-01-2018
30	274	7	57013	83331	2021	PUGA SERRANO ISIDORA TERESA	15.099.261-3	INVERSIONES Y ASESORIAS VALLE ALTO S.A.	05-08-2021	15350		22-07-2021
31	293	22	54797	31402	1996	INMOBILIARIA ISIDORA 2900 SPA	77.066.925-1	INMOBILIARIA NUEVA ISIDORA SPA	18-10-2019	63517.1136		01-10-2019
32	293	23	7308	7837	1997	INMOBILIARIA ISIDORA 2900 SPA	77.066.925-1	INMOBILIARIA NUEVA ISIDORA SPA	28-10-2019	92629.124		01-10-2019
33	293	24	001036	1194	2002	INMOBILIARIA ISIDORA 2900 SPA	77.066.925-1	INMOBILIARIA NUEVA ISIDORA SPA	28-10-2019	108508.4024		01-10-2019

ANEXO 3. RESULTADO PAQUETE DE CARET

only 20 most important variables shown (out of 92)

	Overall
logcoef_de_constructibilidad	7.393
coef_de_constructibilidad	6.840
cuadcoef_de_constructibilidad	6.060
logn_de_pisos	5.729
lograsantes	5.241
cuads_edif_util_total	5.058
altura_maxima	5.033
cuadaltura_maxima	4.980
logaños	4.780
estacionamientos	4.626
cuadrasantes	4.571
rasantes	4.561
cuadn_de_pisos	4.531
n_de_pisos	4.460
cuadaños	4.329
logsuperficie_edificada_m2	4.245
s_edif_util_total	3.973
años	3.837
cuadsuperficie_edificada_m2	3.651
logsuperficie_terreno	3.569

ANEXO 4. PRIMERA REGRESIÓN LINEAL MÚLTIPLE

```

Residuals:
    Min       1Q   Median       3Q      Max
-3.9399 -0.2089  0.0909  0.3499  3.1109

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.875e+02  5.852e+01  -3.205  0.001681 **
cuadcoef_de_constructibilidad  8.733e-02  2.671e-02   3.269  0.001362 **
cuadn_de_pisos    8.703e-04  7.544e-04   1.154  0.250626
años             9.348e-02  2.899e-02   3.225  0.001575 **
cuadsuperficie_edificada_m2 -3.318e-12  5.353e-11  -0.062  0.950666
densidad         2.832e-03  2.087e-03   1.357  0.177062
estacionamientos -1.343e-03  1.130e-03  -1.188  0.236796
cuadsuperficie_terreno -1.336e-10  7.731e-11  -1.728  0.086302 .
rasantes         4.408e-02  1.367e-02   3.224  0.001576 **
s_edif_util_total -4.194e-01  4.853e-01  -0.864  0.388978
cuadaltura_maxima -6.415e-05  2.202e-04  -0.291  0.771279
N_Edificación    7.937e-03  2.083e-02   0.381  0.703752
coef_ocupacion_suelo -1.336e+00  9.201e-01  -1.452  0.148821
viviendas        -5.553e-04  2.037e-03  -0.273  0.785540
locales_comerciales -1.474e-02  1.653e-02  -0.892  0.373909
oficinas         -4.414e-03  7.520e-03  -0.587  0.558210
censal_zona13_6  -2.385e+00  6.649e-01  -3.587  0.000464 ***
censal_zona16_1  -8.245e-01  3.620e-01  -2.277  0.024299 *
censal_zona15_1  -1.135e+00  3.789e-01  -2.996  0.003244 **
censal_zona7_2  -1.022e+00  6.082e-01  -1.681  0.095065 .
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8436 on 138 degrees of freedom
Multiple R-squared:  0.5896,    Adjusted R-squared:  0.5331
F-statistic: 10.44 on 19 and 138 DF,  p-value: < 2.2e-16

```

ANEXO 5. REGRESIÓN LINEAL DESPUÉS DE ELIMINAR VARIABLES INDEPENDIENTES

```

Residuals:
    Min       1Q   Median       3Q      Max
-3.8424 -0.1512  0.0930  0.3654  3.1445

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.951e+02  5.751e+01  -3.392  0.000898 ***
cuadcoef_de_constructibilidad  7.454e-02  2.146e-02   3.474  0.000681 ***
cuadn_de_pisos    7.471e-04  5.160e-04   1.448  0.149865
años             9.695e-02  2.851e-02   3.400  0.000876 ***
cuadsuperficie_edificada_m2  6.909e-12  4.531e-11   0.152  0.879018
densidad        2.966e-03  1.966e-03   1.509  0.133611
estacionamientos -1.481e-03  1.020e-03  -1.451  0.149009
cuadsuperficie_terreno -1.410e-10  6.591e-11  -2.140  0.034100 *
rasantes        4.860e-02  1.185e-02   4.100  6.91e-05 ***
N_Edificación   8.327e-03  2.043e-02   0.408  0.684172
coef_ocupacion_suelo -1.483e+00  8.354e-01  -1.775  0.078065 .
locales_comerciales -2.134e-02  1.260e-02  -1.694  0.092537 .
censal_zona13_6  -2.399e+00  6.580e-01  -3.646  0.000373 ***
censal_zona16_1  -8.166e-01  3.565e-01  -2.290  0.023479 *
censal_zona15_1  -1.229e+00  3.581e-01  -3.433  0.000783 ***
censal_zona7_2   -1.045e+00  6.017e-01  -1.736  0.084695 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8365 on 142 degrees of freedom
Multiple R-squared:  0.5848,    Adjusted R-squared:  0.541
F-statistic: 13.34 on 15 and 142 DF,  p-value: < 2.2e-16

```

ANEXO 6. REGRESIÓN LINEAL CON VARIABLES TRANSFORMADAS

```

Residuals:
    Min       1Q   Median       3Q      Max
-3.4467 -0.2179  0.0620  0.4032  2.9632

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.103e+02  3.576e+01  -3.085  0.00244 **
cuadcoef_de_constructibilidad  4.819e-02  1.859e-02   2.592  0.01055 *
logn_de_pisos    6.834e-01  2.553e-01   2.677  0.00831 **
cuadaños        2.763e-05  8.753e-06   3.156  0.00195 **
logSuperficie_edificada_m2  -2.926e-01  1.706e-01  -1.715  0.08855 .
densidad        5.380e-03  1.803e-03   2.985  0.00334 **
estacionamientos -9.836e-04  3.368e-04  -2.920  0.00407 **
superficie_terreno -3.377e-05  5.933e-06  -5.692  6.94e-08 ***
cuadrasantes    3.425e-04  1.296e-04   2.643  0.00913 **
N_Edificación   4.760e-02  2.190e-02   2.173  0.03142 *
logcoef_ocupacion_suelo -6.828e-01  2.926e-01  -2.334  0.02103 *
locales_comerciales -2.055e-02  1.325e-02  -1.551  0.12304
censal_zona13_6  -1.981e+00  2.244e+00  -0.883  0.37893
censal_zona16_1  -8.795e-01  6.958e-01  -1.264  0.20826
censal_zona15_1  -9.942e-01  3.087e-01  -3.220  0.00159 **
censal_zona7_2   -1.044e+00  8.999e-01  -1.160  0.24804
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7931 on 142 degrees of freedom
Multiple R-squared:  0.6267,    Adjusted R-squared:  0.5873
F-statistic: 30.76 on 15 and 142 DF,  p-value: < 2.2e-16

```


ANEXO 6. REGRESIÓN LINEAL MÉTODO ESTIMACIÓN ROBUSTOS

```
Call:
lm(formula = logvalorm2 ~ cuadcoef_de_constructibilidad + logn_de_pisos +
  cuadaños + logsuperficie_edificada_m2 + densidad + estacionamientos +
  superficie_terreno + quadrasantes + N_Edificación + logcoef_ocupacion_su
  locales_comerciales + censal_zona13_6 + censal_zona16_1 +
  censal_zona15_1 + censal_zona7_2, data = datos)
```

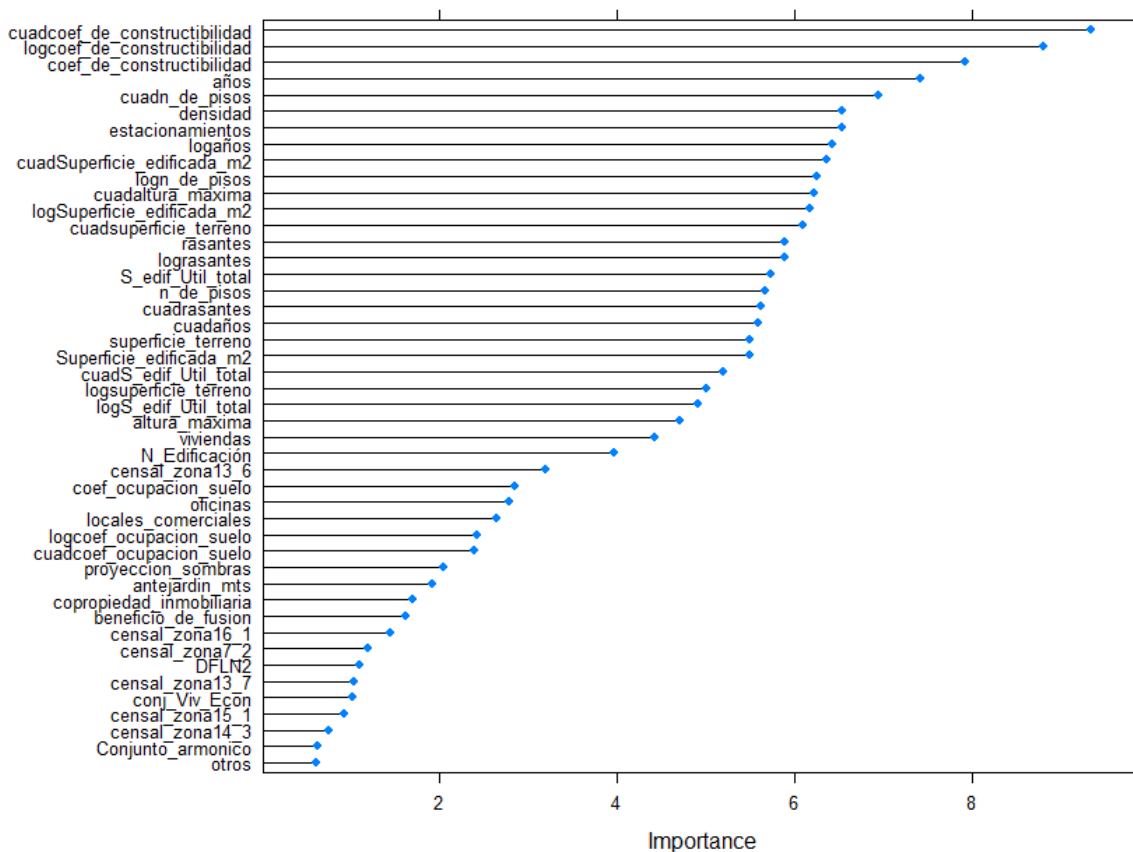
```
Residuals:
    Min       1Q   Median       3Q      Max
-3.4467 -0.2179  0.0620  0.4032  2.9632
```

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -1.103e+02  2.861e+01  -3.856 0.000174 ***
cuadcoef_de_constructibilidad  4.819e-02  2.119e-02   2.274 0.024469 *
logn_de_pisos      6.834e-01  2.086e-01   3.276 0.001324 **
cuadaños          2.763e-05  7.003e-06   3.945 0.000125 ***
logsuperficie_edificada_m2  -2.926e-01  1.180e-01  -2.479 0.014347 *
densidad          5.380e-03  1.743e-03   3.086 0.002439 **
estacionamientos  -9.836e-04  4.076e-04  -2.413 0.017090 *
superficie_terreno -3.377e-05  9.046e-06  -3.733 0.000273 ***
quadrasantes      3.425e-04  1.165e-04   2.940 0.003828 **
N_Edificación     4.760e-02  2.152e-02   2.212 0.028570 *
logcoef_ocupacion_suelo  -6.828e-01  2.546e-01  -2.682 0.008189 **
locales_comerciales -2.055e-02  1.039e-02  -1.977 0.049935 *
censal_zona13_6   -1.981e+00  6.398e-01  -3.096 0.002362 **
censal_zona16_1   -8.795e-01  3.397e-01  -2.590 0.010611 *
censal_zona15_1   -9.942e-01  3.413e-01  -2.913 0.004157 **
censal_zona7_2    -1.044e+00  5.710e-01  -1.828 0.069646 .
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 0.7931 on 142 degrees of freedom
Multiple R-squared:  0.6267,    Adjusted R-squared:  0.5873
F-statistic: 15.9 on 15 and 142 DF,  p-value: < 2.2e-16
```

ANEXO 7. RESULTADO CARET CON VARIABLES ELEGIDAS

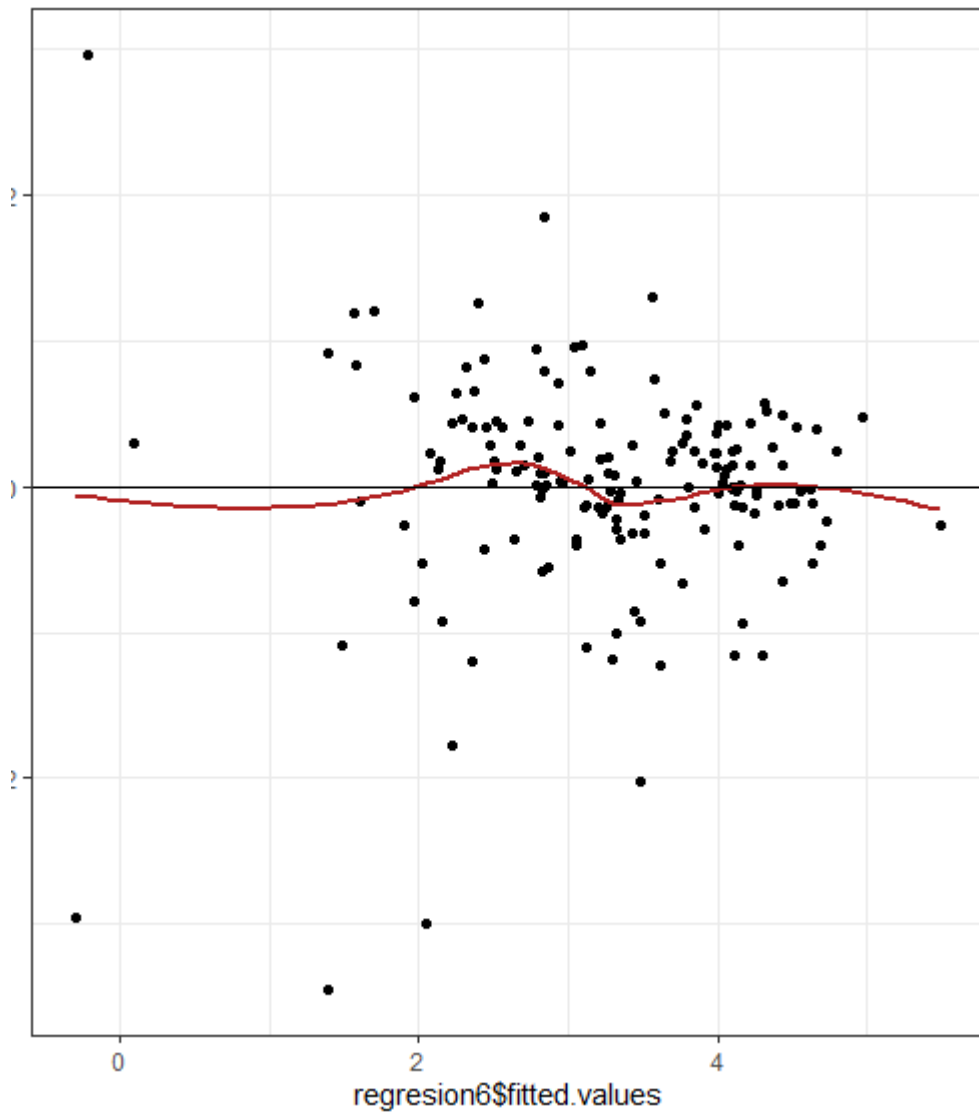


ANEXO 8. PRUEBA DURBIN-WATSON

Durbin-watson test

```
data: regresion6
Dw = 1.9593, p-value = 0.373
```

ANEXO 9. RESULTADO RESIDUOS REGRESIÓN LINEAL



ANEXO 10. DESVIACIÓN ESTÁNDAR COEFICIENTE DE CONSTRUCTIBILIDAD

```
> summary(desvest)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
1.233  1.233  1.233  1.233  1.233  1.233
> |
```

ANEXO 11. DESVIACIÓN ESTÁNDAR DENSIDAD

```
> summary(desvest)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
95.25  95.25  95.25  95.25  95.25  95.25
.....
```

ANEXO 12. COMPARACIÓN VALORES REALES VS VALORES PREDICTIVOS DE LA REGRESIÓN LINEAL

n° de muestras	resultados regresión lineal	valores reales del suelo	diferencia absoluta de valor real vs predicción
1	1,9710	2,58	0,6089906
2	2,1258	2,24	0,1141972
3	2,0831	2,32	0,2368869
4	2,1438	2,32	0,1762365
5	3,5168	3,32	0,1967667
6	4,2143	4,36	0,1456745
7	3,0133	3,26	0,2466953
8	4,3616	4,64	0,2783919
9	2,4396	3,32	0,8804175
10	3,4256	3,72	0,2944285
11	3,5085	3,19	0,3185106
12	2,8310	2,25	0,580984
13	5,4876	5,22	0,2676469
14	2,8078	2,74	0,0677849
Promedio	3,1378	3,2486	0,3153