

## Avance 1 proyecto

**Elaborado por:** Andres Santiago Serna Tangarife, Katherine Russi Parra, William Morales y Daniela Castillo Téllez

### Problema que abordarán y su contexto.

El proyecto se centra en el desarrollo de un modelo de aprendizaje automático para predecir los precios de vivienda en Colombia. Esta iniciativa surge en un entorno en el que los compradores y vendedores de propiedades inmobiliarias necesitan información precisa y confiable para tomar decisiones informadas en el mercado inmobiliario. Dado que los precios de las viviendas pueden fluctuar considerablemente en un país como Colombia, es crucial contar con un modelo predictivo que pueda proporcionar estimaciones precisas y oportunas de los precios de las propiedades en diferentes áreas geográficas.

El conjunto de datos recopilados incluye información detallada sobre una variedad de propiedades en diferentes departamentos y ciudades de Colombia, lo que proporciona una amplia cobertura geográfica. La recopilación de datos se llevó a cabo a través de un proveedor de datos y técnicas de scraping de diversas páginas web de venta de inmuebles.

La iniciativa también tiene como objetivo mantener la viabilidad a largo plazo del modelo, con la implementación de un plan para la recopilación periódica de datos que permitirá reentrenar el modelo y validar su precisión continua.

### Pregunta de negocio y alcance del proyecto.

La pregunta de negocio central que aborda este proyecto es:

¿Cómo predecir con precisión los precios de vivienda en Colombia para facilitar a compradores y vendedores la toma de decisiones informadas en el mercado inmobiliario?

El alcance del proyecto abarca varios aspectos clave:

- **Desarrollo de un modelo de aprendizaje automático:** Se busca crear un modelo robusto y preciso que pueda predecir los precios de vivienda en Colombia utilizando datos detallados de propiedades, ubicación, características y condiciones.
- **Recopilación y almacenamiento de datos:** Se ha recopilado y almacenado un conjunto inicial de datos de 12,853 propiedades de diferentes departamentos y ciudades de Colombia. Se establece un plan para la recolección periódica de datos con el fin de reentrenar el modelo y validar su precisión a lo largo del tiempo.
- **Despliegue de un tablero de control en la nube:** Se prevé implementar un tablero de control accesible para los usuarios, que permita un análisis descriptivo inicial y el acceso a las predicciones generadas por el modelo. El tablero brindará una interfaz intuitiva para que los usuarios comprendan y utilicen los resultados del modelo.

### Conjuntos de datos a emplear

Los datos a emplear corresponden a datos de 12.853 inmuebles de diferentes departamentos y ciudades de Colombia. El formato de los datos es un archivo de Excel donde se cuenta con toda la información de las propiedades y sus respectivos avalúos.

La recolección de datos se realizó a partir de un proveedor de datos con la información de diferentes proyectos de vivienda en Colombia cuya información se complementa utilizando técnicas de scraping de diferentes páginas web de venta de inmuebles en Colombia.

El objetivo de estos datos es estimar el precio total de un inmueble dadas unas características de este. En el conjunto de datos existe una columna llamada `valor_total_avaluo`, la cual es una variable numérica, que indica el precio total del inmueble que incluye áreas privadas, públicas, parqueaderos y cuarto útil en caso de tener alguno de estos.

Dentro del conjunto de datos se cuenta con los siguientes grupos de variables:

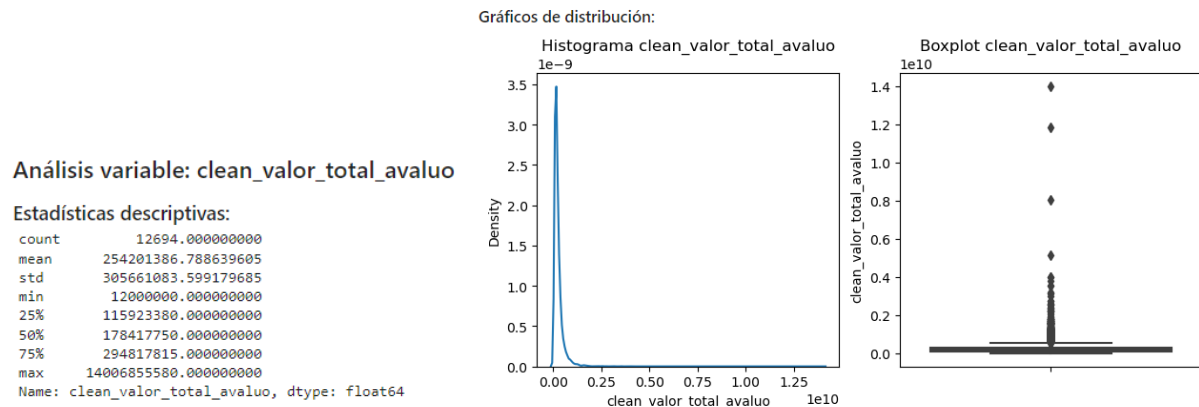
Grupo de Variables	Descripción	Variables
Datos Generales	Este grupo incluye información básica sobre la ubicación y características generales del inmueble.	Motivo, Departamento_inmueble, Municipio_inmueble, Barrio, Direccion_inmueble_informe, Estrato, Topografia_sector, Nivel_equipamiento_comercial, Vetustez, Longitud, Latitud, Clean_valor_total_avaluo
Características Básicas	Estas variables proporcionan información sobre el tipo de inmueble y su estructura.	Tipo_inmueble, Clase_inmueble, Ocupante, Sometido_a_propiedad_horizontal, Altura_permitida, Numero_piso, Numero_de_edificios, Numero_total_de_garajes, Total_cupos_parquedero, Numero_total_depositos, Area_libre
Infraestructura Vial	Este grupo describe las condiciones de las vías cercanas al inmueble.	Vias_pavimentadas, Sardineles_en_las_vias, Andenes_en_las_vias, Paradero, Arborizacion, Alamedas, Ciclo_rutas
Servicios Públicos	Estas variables indican la disponibilidad de servicios públicos en el inmueble.	Gas_en_el_predio, Telefono_en_el_predio
Instalaciones Comunitarias	Este grupo describe las instalaciones compartidas dentro del conjunto residencial o edificio.	Porteria, Citofono, Bicicletero, Piscina, Tanque_de_agua, Club_house, Teatrino, Sauna, Vigilancia_privada, Tipo_vigilancia, Administracion
Estructura del Inmueble	Estas variables proporcionan información sobre la estructura física del inmueble.	Pisos_bodega, Estructura, Ajustes_sismoresistentes, Cubierta, Fachada, Tipo_fachada, Estructura_reforzada, Danos_previos, Material_de_construccion, Detalle_material
Distribución Interna	Este grupo describe la distribución interna del inmueble.	Habitaciones, Estar_habitacion, Cuarto_servicio, Closet, Sala, Comedor, Bano_privado, Bano_social, Bano_servicio, Cocina, Estudio, Balcon, Terraza, Patio_interior, Jardin, Zona_de_ropas
Acabados	Estas variables detallan el estado y calidad de los acabados en el inmueble.	Estado_acabados_pisos, Calidad_acabados_pisos, Calidad_acabados_muros, Calidad_acabados_techos, Estado_acabados_madera, Calidad_acabados_madera, Calidad_acabados_metal, Calidad_acabados_banos, Estado_acabados_cocina, Calidad_acabados_cocina
Estacionamiento	Esta categoría indica el tipo de garaje y depósito disponibles en el inmueble	Tipo_garaje, Tipo_deposito
Características Físicas	Estas variables describen características físicas específicas del inmueble.	Area_valorada, Indice_ocupacion, Indice_construccion, Aislamiento_posterior, Aislamiento_lateral, Antejardin, Condicion_ph, Rph, Area_libre, Estado_acabados_pisos, Calidad_acabados_pisos, Calidad_acabados_muros, Calidad_acabados_techos, Estado_acabados_madera, Calidad_acabados_madera, Calidad_acabados_metal, Calidad_acabados_banos, Estado_acabados_cocina, Calidad_acabados_cocina, Irregularidad_planta, Irregularidad_altura

**Repositorio Git en uso para el código:** [https://github.com/asernat/miad\\_proyecto\\_despliegue](https://github.com/asernat/miad_proyecto_despliegue)

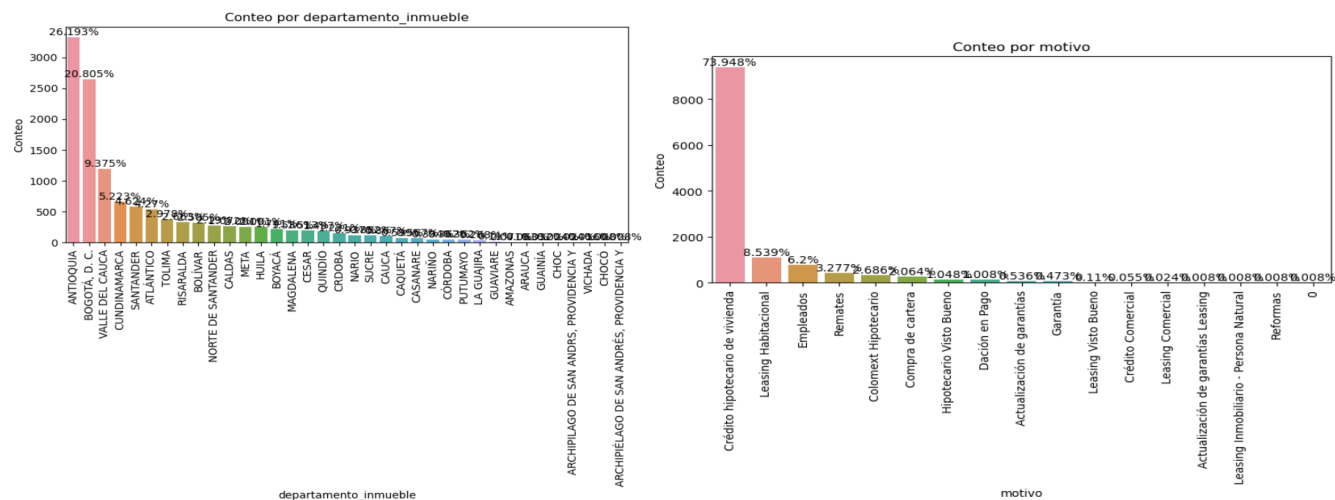
**Repositorio DVC en uso para los datos:** El repositorio DVC para los datos se realizó en Google Drive en la siguiente carpeta: <https://drive.google.com/drive/folders/1HyhWYmunC9PaANnc6PIHug9vJFGUKrrA>

## Exploración de los datos.

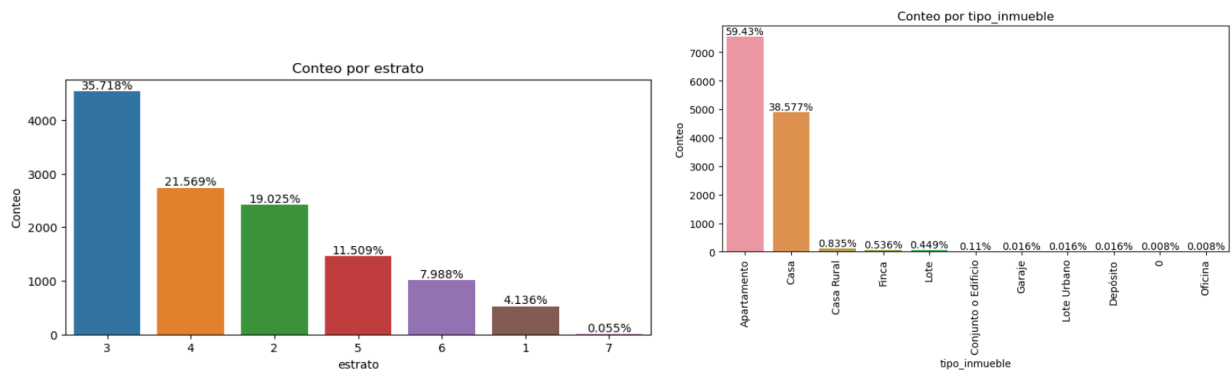
Al realizar el análisis exploratorio de los datos, principalmente en los valores de venta de inmuebles, se puede observar que el valor mínimo es 12.000.000, el valor máximo es 14.006.855.580 y que la mayoría de los datos se encuentran en el rango entre 115.923.380 y 294.817.815.



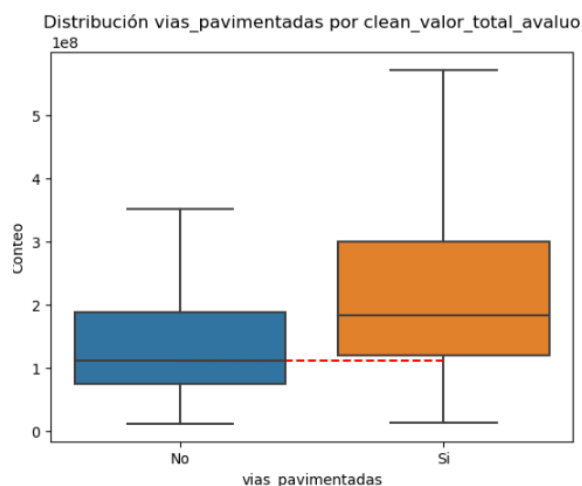
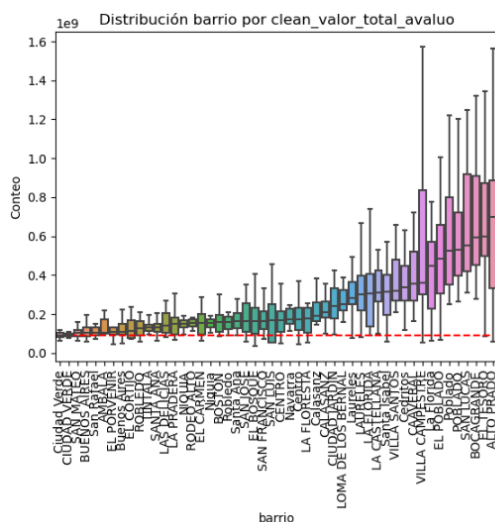
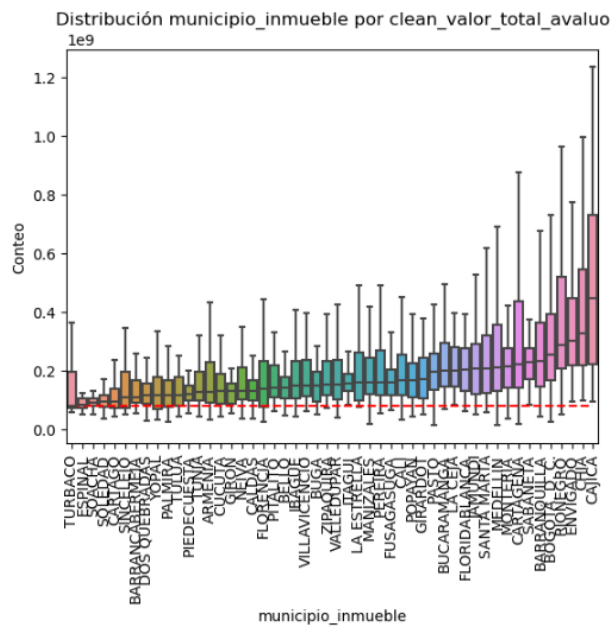
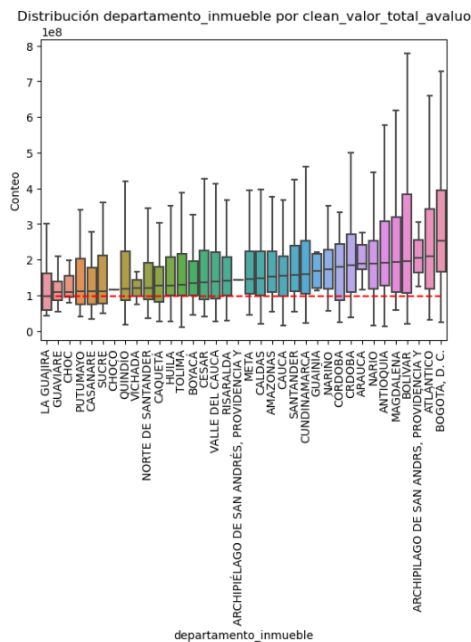
Sobre la ubicación de los diferentes inmuebles con los que se cuentan dentro de la base de datos, se puede evidenciar que alrededor del 60% de los inmuebles se encuentran en Antioquia, Bogotá, Valle del Cauca y Cundinamarca. Por otro lado, los principales motivos de compra se dan por crédito hipotecario de vivienda y leasing habitacional que sumado representan el 82,4% dentro de la base.



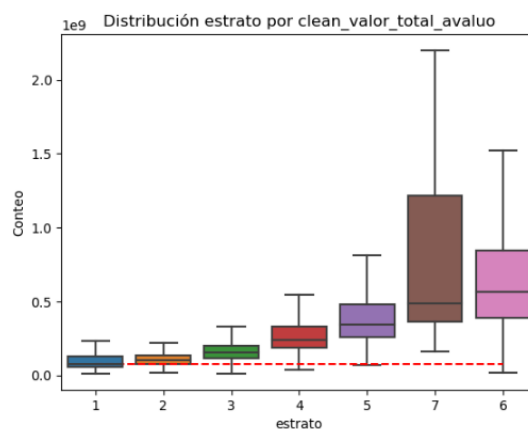
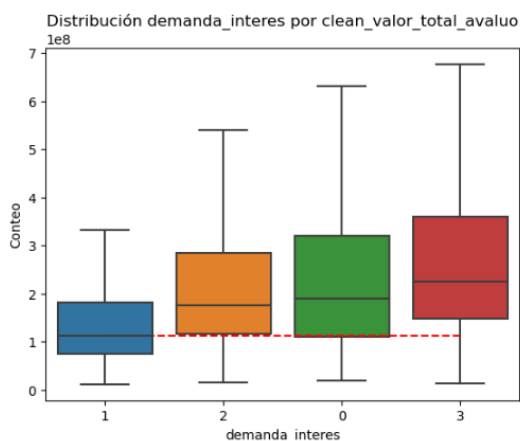
La mayoría de los inmuebles en la base de datos pertenecen a los estratos 3, 4 y 2, siendo el estrato 3 el más predominante con un 35.71% y de forma transversal la base se encuentra compuesta en su mayoría por casas y apartamentos.

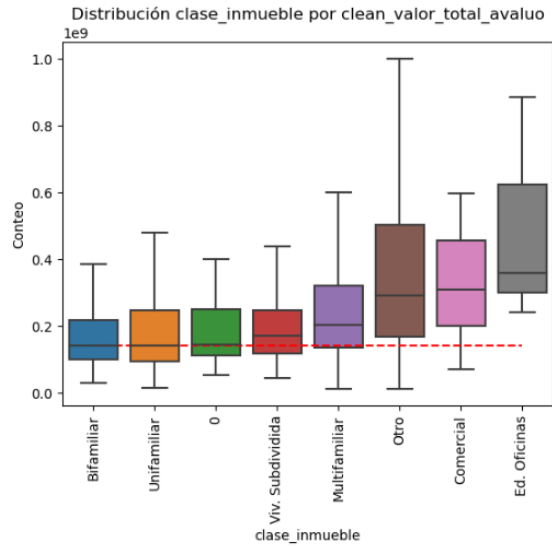
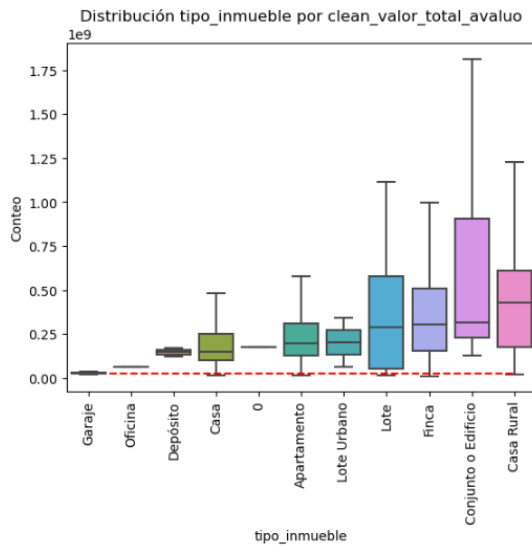


Dentro de la base de datos, hemos identificado la influencia de diversas características de los inmuebles en sus valores de venta. Entre estas características se incluyen la ubicación geográfica, especificada por departamento, municipio y barrio, así como la disponibilidad de acceso a través de vías pavimentadas.

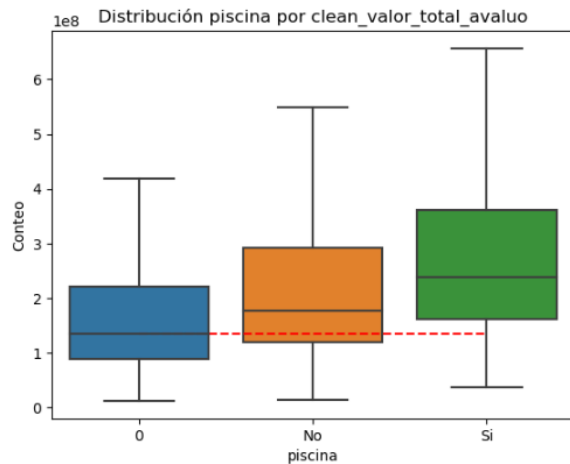
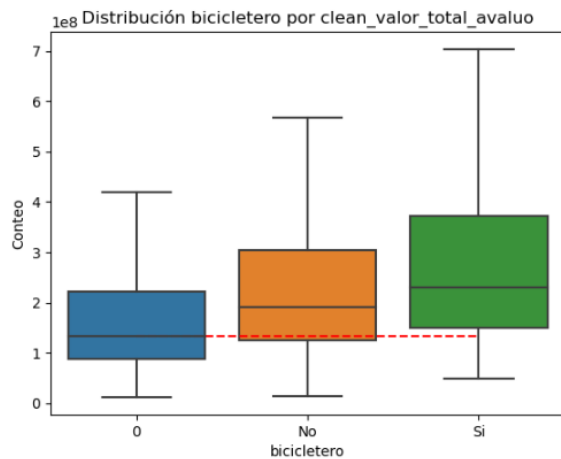
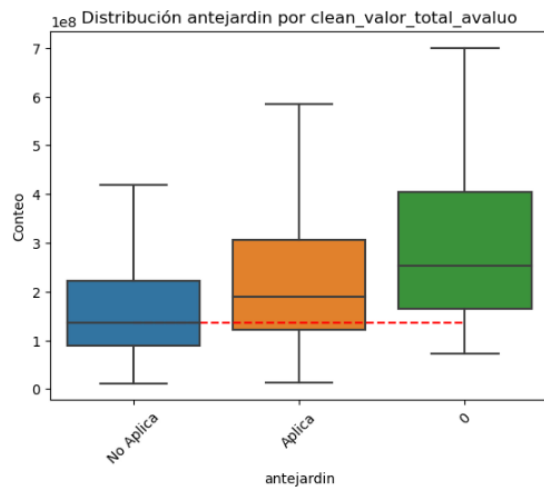
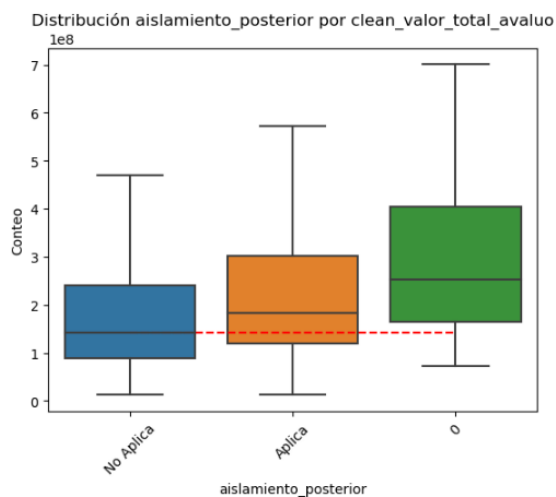


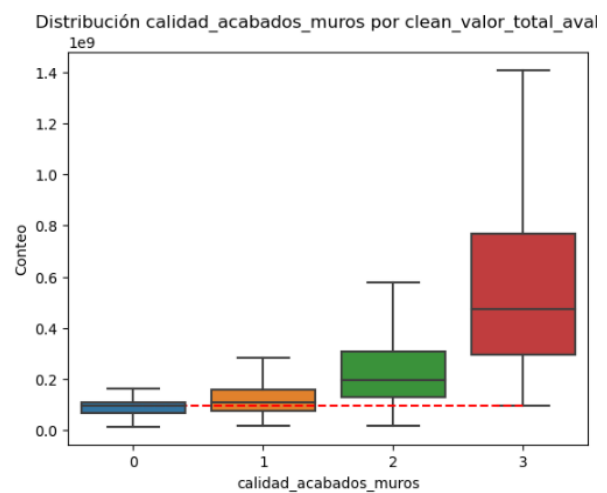
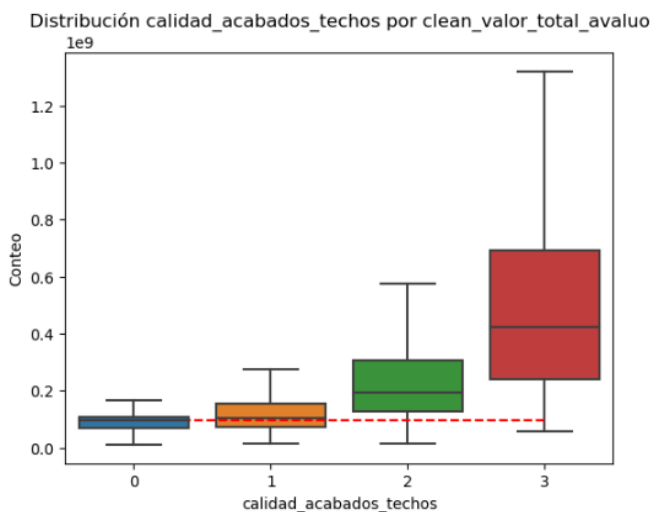
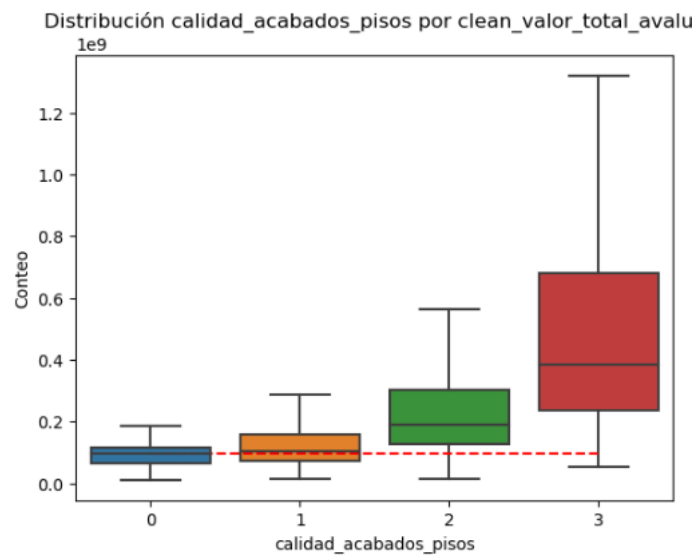
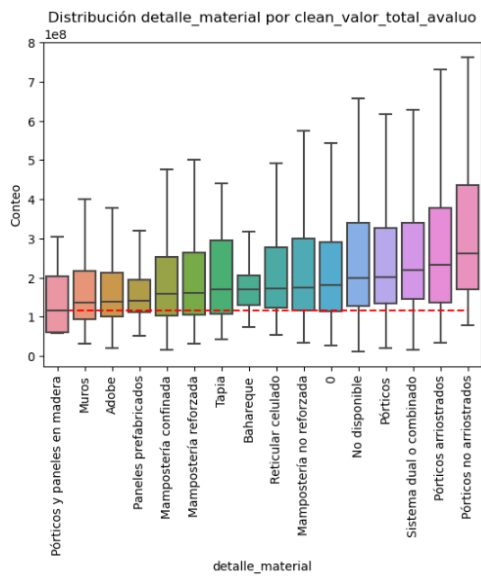
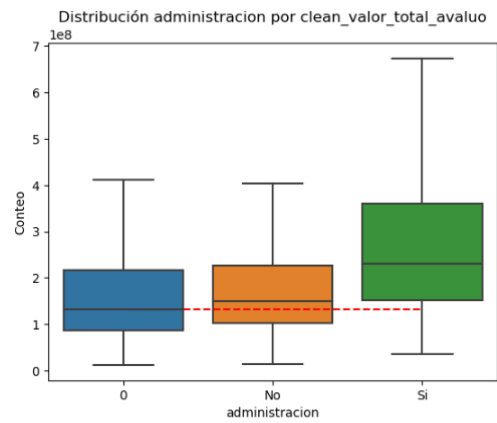
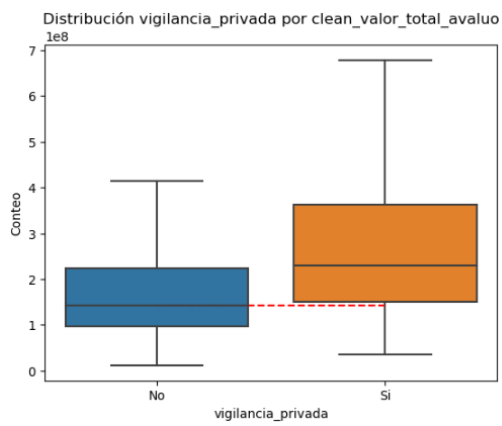
A pesar de que el estrato tiene una influencia significativa en el valor de venta del inmueble, también es esencial considerar la clase y tipo de inmueble. Estas variables son fundamentales, ya que determinan si estamos tratando con propiedades de uso comercial o familiar, así como si se trata de conjuntos residenciales, casas en áreas urbanas o rurales.



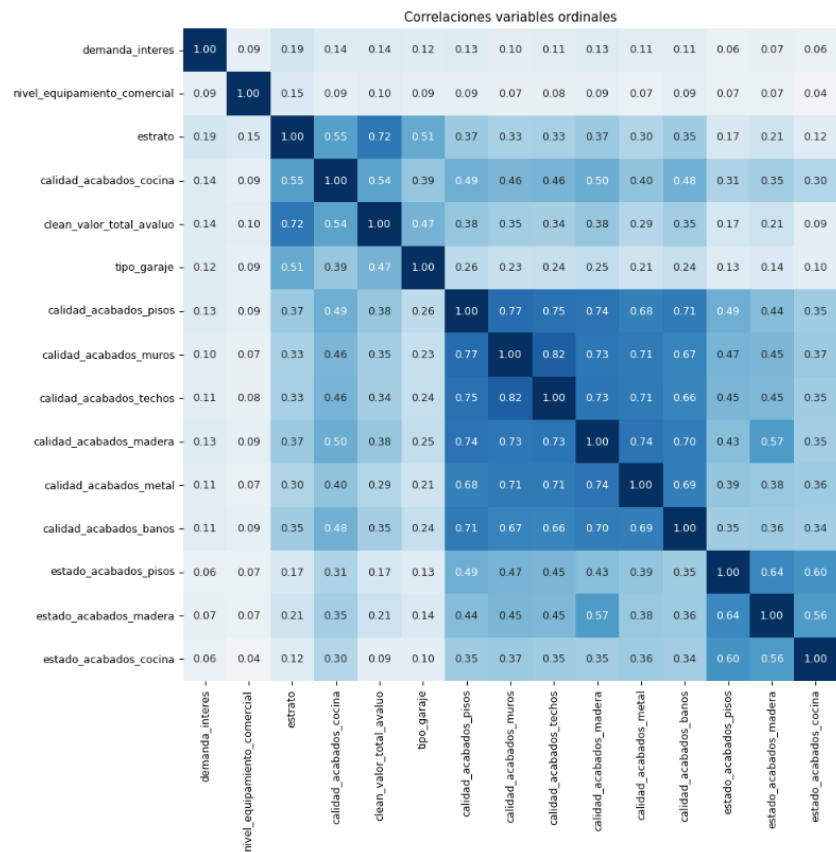


La condición física del inmueble y sus complementos, como piscinas, áreas para bicicletas, jardines delanteros, entre otros, desempeñan un papel importante al momento de determinar su valor. Además, la calidad de los acabados y características también son factores fundamentales. Aspectos como la calidad de los pisos, muros y techos, incluyendo los detalles de los materiales empleados en la construcción, tienen una influencia significativa en la percepción de calidad y, por ende, en el valor final del inmueble.

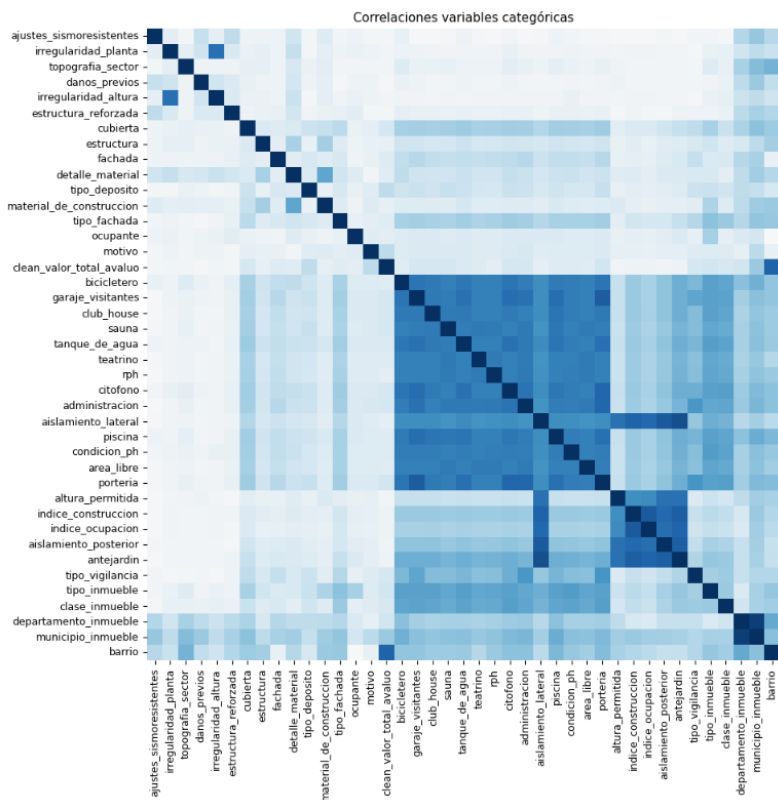




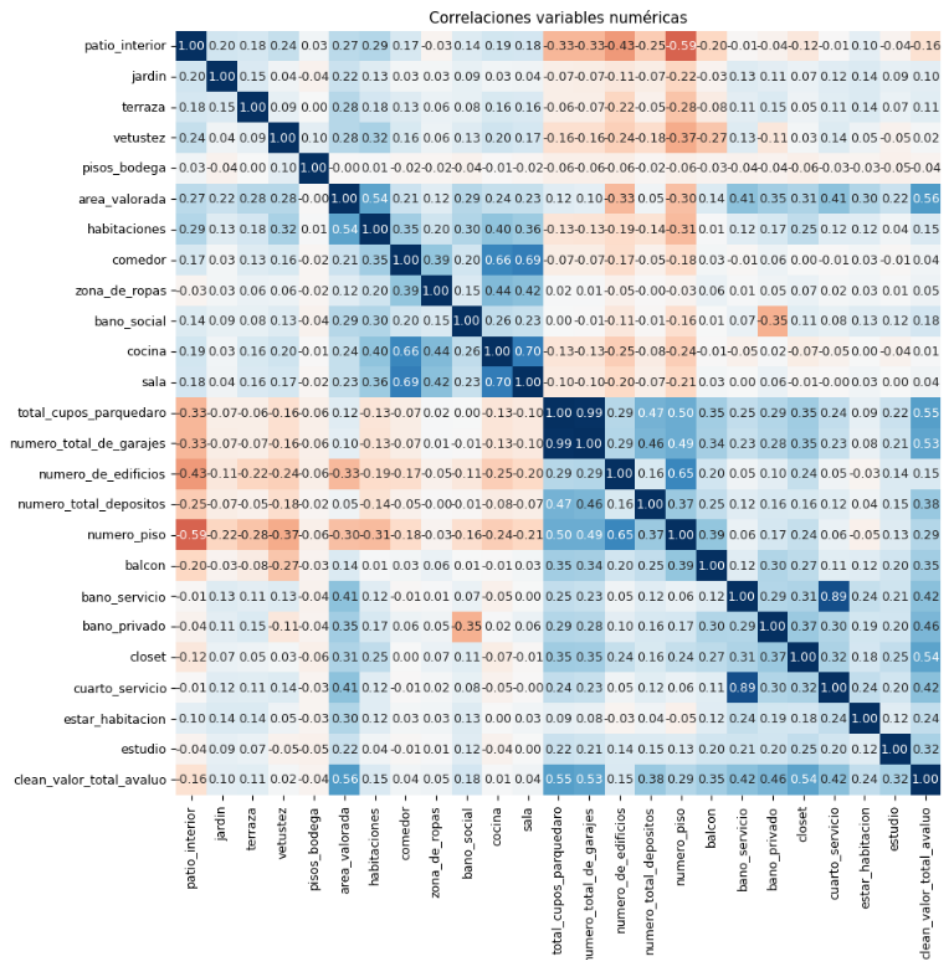
A nivel de correlación de variables ordinales, se observa principalmente una alta correlación asociada a la calidad de los acabados.



En términos de variables categóricas, que en su mayoría en la base de datos están representadas como respuestas binarias (sí/no), se destaca una correlación significativa entre las características técnicas del inmueble y su ubicación geográfica, incluyendo el departamento, municipio y barrio. Estas relaciones resaltan la importancia de entender cómo las especificaciones técnicas y la ubicación influyen en el valor de venta.



Con la variable objetivo del caso que es el valor de venta del inmueble, muestra una correlación significativa, especialmente en relación con la ubicación y los acabados del inmueble.



Top 10 variables más correlacionadas con variable objetivo

	index	variable	value
3080	clean_valor_total_avaluo	barrio	0.809693882
3238	clean_valor_total_avaluo	estrato	0.724232808
4344	clean_valor_total_avaluo	area_valorada	0.564292638
6003	clean_valor_total_avaluo	total_cupos_parquedaro	0.547982730
4897	clean_valor_total_avaluo	closet	0.536990330
4186	clean_valor_total_avaluo	calidad_acabados_cocina	0.535568035
5924	clean_valor_total_avaluo	numero_total_de_garajes	0.528678813
4265	clean_valor_total_avaluo	tipo_garaje	0.474271510
5134	clean_valor_total_avaluo	banoprivado	0.457182334
4818	clean_valor_total_avaluo	cuarto_servicio	0.420292650

**Nota:** la base de datos principal con la que trabajamos se encuentra en el dvc sin embargo [acá](#) encontrarán el excel de dicha información inicial



- Maqueta del prototipo.

## PESTAÑA 1: navegar la ubicación de las propiedades y visualizar sus características

**Comprar o vender una vivienda en Colombia**

Este sitio le permite obtener el precio justo de vivienda para compra o venta en Colombia relacionada a sus preferencias

**Total Inmuebles:**

Cantidad !

**Total Inmuebles:**

**Precio promedio vivienda**

Cantidad !

Área(m2):

Text input

**Seleccione el filtro:**

Seleccione Estado: Item 1, Item 2, Item 3

Seleccione Departamento: Item 1, Item 2, Item 3

Seleccione Municipio: Item 1, Item 2, Item 3

Seleccione Barrio: Item 1, Item 2, Item 3

Seleccione Tipo Inmueble: Item 1, Item 2, Item 3

**Mapa de Colombia Cantidad de inmuebles**

**Cantidad de viviendas por tipo de Inmueble:**

**Precio promedio viviendas por tipo de Inmueble:**

## PESTAÑA 2: para ingresar datos y obtener la predicción del precio

**Predecir precios de vivienda en Colombia**

Seleccione las características para realizar la predicción de precio

**Ubicación:**

Seleccione Departamento: Item 1, Item 2, Item 3

Seleccione Municipio: Item 1, Item 2, Item 3

Seleccione Barrio: Item 1, Item 2, Item 3

**Características del Inmueble:**

Seleccione Tipo Inmueble: Item 1, Item 2, Item 3

Seleccione Estado: Item 1, Item 2, Item 3

Área(m2):

Text input

Este sitio le permite predecir los precios de vivienda para compra o venta en Colombia, contiene información de gran variedad de propiedades y amplia cobertura geográfica del País.

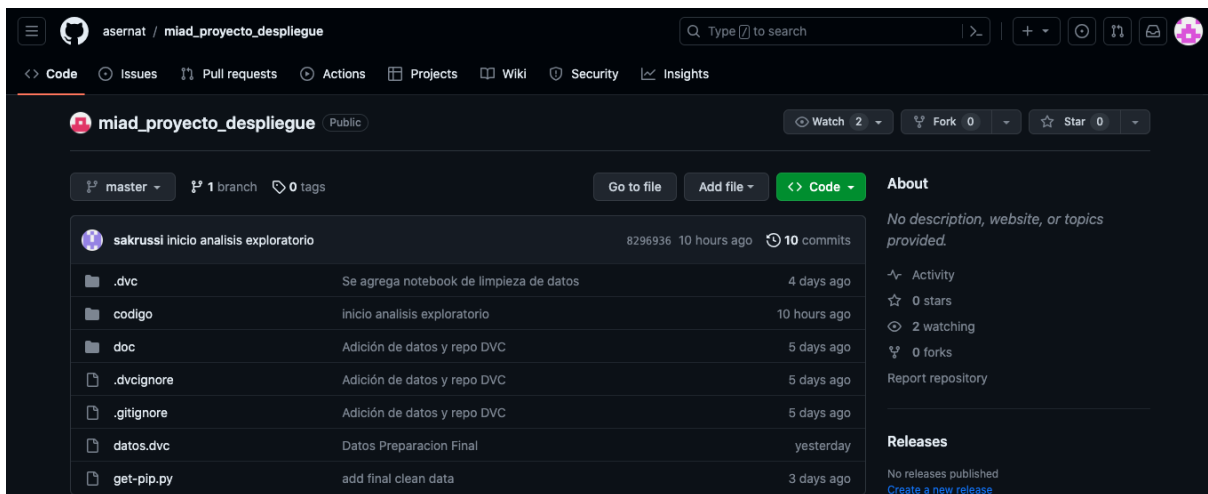
**Precio de vivienda:**

Resultado Predicción

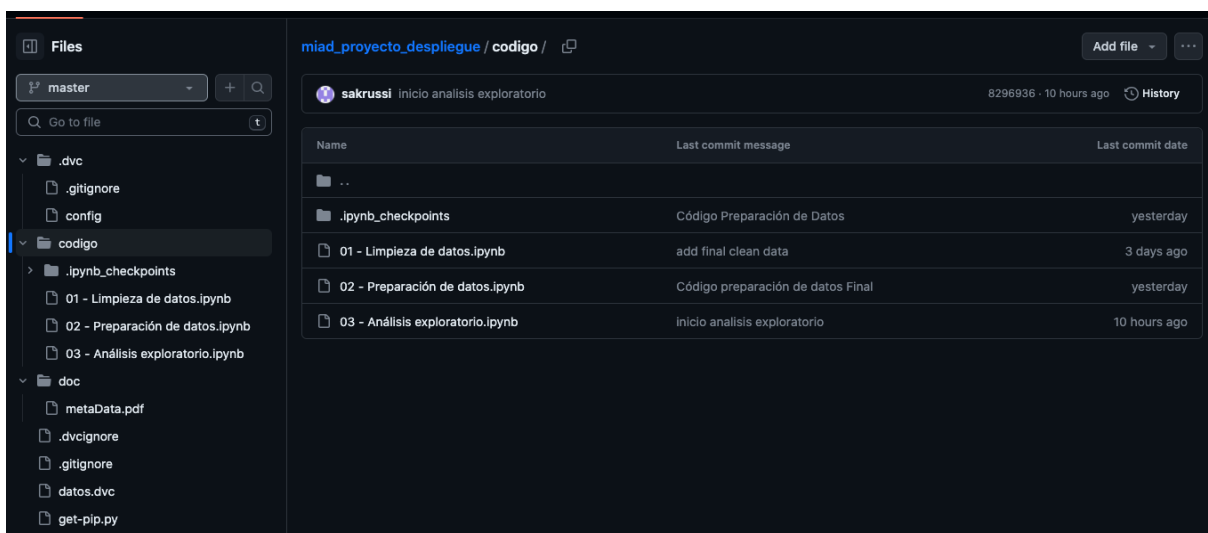
- Repositorios con todo el código (capturas y enlaces que soporten la creación y uso de repositorios).

[https://github.com/asernat/miad\\_proyecto\\_despliegue](https://github.com/asernat/miad_proyecto_despliegue)

## 1. creación de repositorio Github

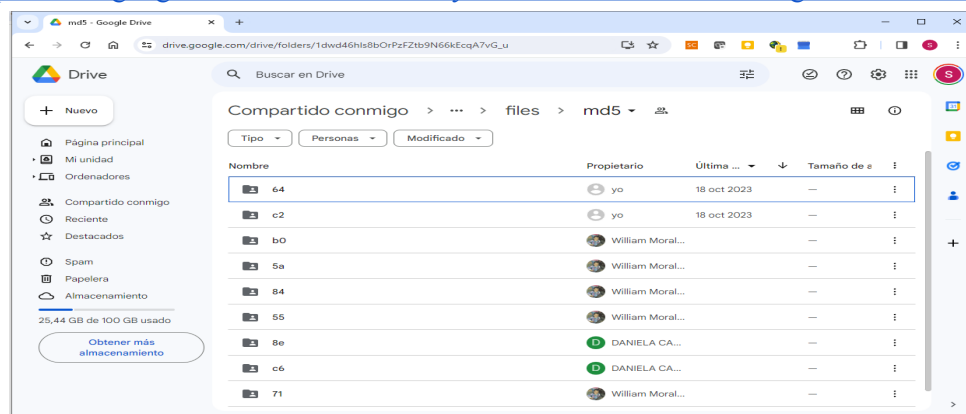


## 2. carga de notebooks al repositorio por cada integrante:



- Repositorio DVC

- <https://drive.google.com/drive/folders/1HyhWYmunC9PaANnc6PIHug9vJFGUKrrA?usp=sharing>



- **Reporte de trabajo en equipo**

Nombre	Responsabilidades	Commits
Todos	<ul style="list-style-type: none"> <li>- Problema que abordarán y su contexto</li> <li>- Pregunta de negocio y alcance del proyecto.</li> </ul>	
Andres Santiago Serna Tangarife	<ul style="list-style-type: none"> <li>- Revisión y disponibilidad de fuentes de datos</li> </ul>	
Andres Santiago Serna Tangarife	<ul style="list-style-type: none"> <li>- Descripción de conjuntos de datos a emplear - metadata</li> <li>- Repositorio Git en uso para el Código</li> <li>- Repositorio DVC en uso para los datos</li> </ul>	<ul style="list-style-type: none"> <li>- Creación repositorio</li> <li>- <a href="#">Adición de datos y repo DVC</a></li> </ul>
Daniela Castillo	<ul style="list-style-type: none"> <li>- Limpieza de datos</li> </ul>	<ul style="list-style-type: none"> <li>- <a href="#">Se agrega notebook de limpieza de datos</a></li> </ul>
William Morales	<ul style="list-style-type: none"> <li>- Preparación de datos</li> <li>- Descripción de análisis exploratorio</li> </ul>	<ul style="list-style-type: none"> <li>- <a href="#">Preparación de datos</a></li> </ul>
Katherine Russi Parra	<ul style="list-style-type: none"> <li>- Análisis exploratorio</li> <li>- Maqueta prototipo</li> </ul>	<ul style="list-style-type: none"> <li>- <a href="#">inicio analisis exploratorio</a></li> </ul>