Pontificia Universidad JAVERIANA Bogotá

Facultad de Ciencias
Departamento de Matemáticas

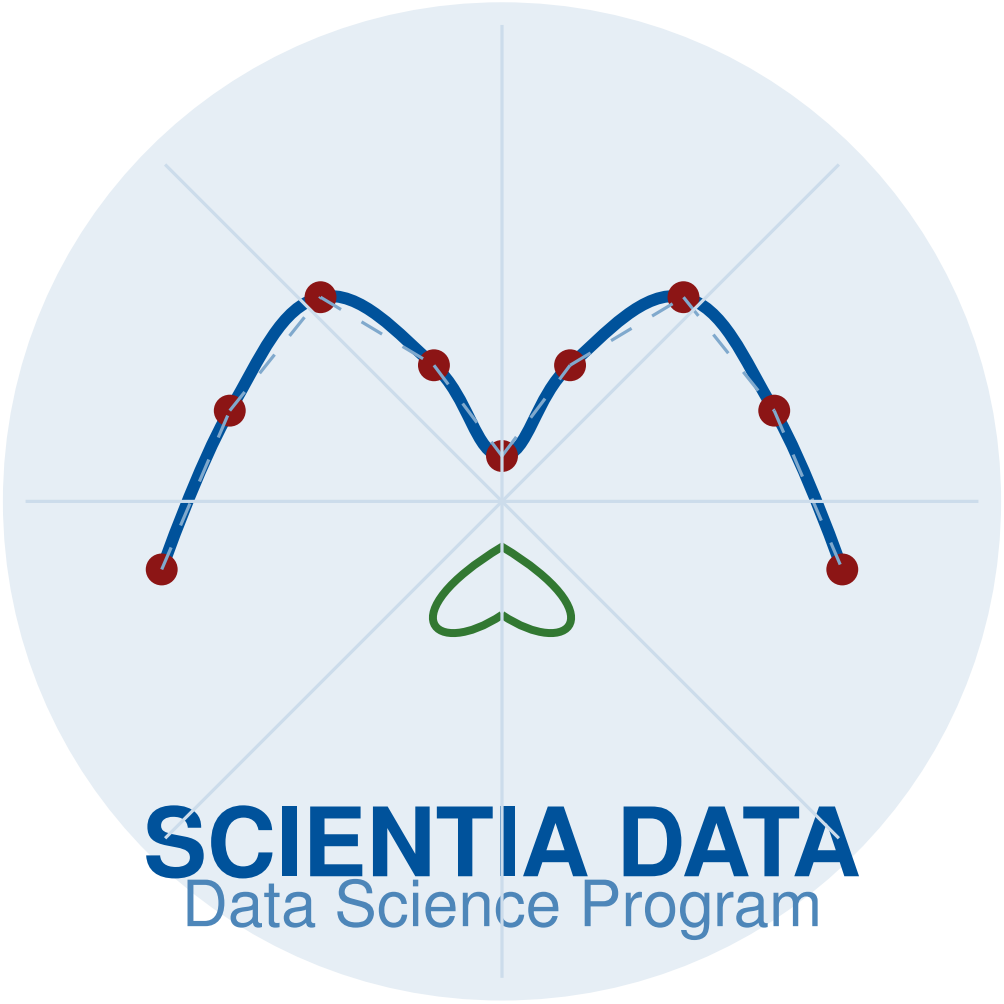
PROGRAMA CIENCIA DE DATOS
TRABAJO DE GRADO

Índice de Precios Inmobiliarios por Bienestar y Entorno Urbano

NEYL PEÑUELA BERNATE

Director: **Vladimir Moreno G.**

11 de junio de 2026
Bogotá, Colombia


SCIENTIA DATA
Data Science Program

Dedicatoria

«Y allí dentro está la voluntad que no muere. ¿Quién conoce los misterios de la voluntad y su fuerza?»

Edgar Allan Poe, Ligeia

Esta fuerza de voluntad de la que tanto hablo no tiene mayor significado que el amor profundo por la familia que me ha tocado. Los amo inmensamente.

Gracias a mi abuelita Esther y a mi abuelito Edgar, por ser ese motor incondicional, por abrazarme incluso cuando sentía no merecerlo, y por esa bondad suya que llena mares.

Gracias a mi mamá Natalia y a mi papá Elkin, por seguirme el juego en todas mis locuras y por acompañarme en mis estrellos.

Gracias a Charlotte y Shadow; he tenido la inmensa suerte de que sean mis compañeros en este tramo de la vida. Los llevo tatuados en el corazón, mis gatos negros hermosos.

Y por último, gracias a mí mismo, por atreverme a creer y a pensar en grande. Y para mi yo del futuro:

«Ante el inmenso oleaje de las probabilidades, jamás reserves nada para el nado de regreso.»

Inspirado en Gattaca (1997)

Resumen | Abstract

RESUMEN

La valoración residencial tradicionalmente se ha centrado en los atributos físicos de los inmuebles y en su precio de mercado, una mirada que ha limitado la posibilidad de comprender de forma objetiva la calidad del entorno urbano en el que esos inmuebles se insertan. Para abordar esa brecha, el presente trabajo propone el desarrollo del Índice Urbano Global (IUG), una herramienta poblacional y descriptiva que sintetiza cinco dimensiones críticas del bienestar urbano en Bogotá: calidad intrínseca, seguridad tridimensional, accesibilidad, proximidad dotacional y potencial normativo. Mediante un enfoque de *Big Data* espacial, se procesó una muestra de 23761 inmuebles georreferenciados, integrando datos de oferta web con información oficial del Distrito (POT 555, IDECA, EPV 2024 e ICSU). Para evaluar la validez de uso del índice, se estimó un modelo hedónico de referencia que, en su mejor especificación funcional, alcanzó un poder explicativo de $R^2 = 0.83$, y se documentó que la incorporación del IUG al modelo agregado eleva el R^2 ajustado en una magnitud estadísticamente significativa bajo bootstrap y validación cruzada. Lejos de buscar redundancia predictiva, el IUG demostró una ortogonalidad deliberada frente al precio ($\rho = -0.181$) y al estrato socioeconómico (AUC $= 0.483$), una independencia confirmada mediante un protocolo de once pruebas estadísticas que incluyen análisis de *commonality*, validación cruzada espacial por localidad, perturbación de pesos por Monte Carlo y bootstrap. Esta independencia estadística permitió cruzar el índice con los residuales del modelo hedónico para detectar ineficiencias de mercado (*mispricing*) mediante una matriz de decisión 2×2 que clasifica cada inmueble en uno de cuatro cuadrantes accionables: comprar, conservar, evitar o vender. El análisis reveló que el 30.7 % de la oferta se ubica en el cuadrante óptimo de inversión, ofreciendo propiedades en promedio 25 % más económicas que las similares pero sobrevaloradas, mientras que un 24.7 % adicional se concentra en el cuadrante de desinversión. Finalmente, el estudio se operativiza a través de una plataforma interactiva equipada con un asistente conversacional basado en arquitectura RAG, democratizando la toma de decisiones inmobiliarias fundamentadas en datos.

Palabras clave: Índice Urbano Global; indicador sintético compuesto; modelo hedónico; análisis geoespacial; búsqueda semántica; calidad urbana; RAG; mercado inmobiliario; Bogotá.

ABSTRACT

Residential valuation has traditionally focused on the physical attributes of properties and their market price, a perspective that has limited the possibility of objectively understanding the quality of the urban environment in which those properties are embedded. To address this gap, this work proposes the development of the Urban Global Index (IUG), a population-level and descriptive tool that synthesizes five critical dimensions of urban well-being in Bogotá: intrinsic quality, three-dimensional security, accessibility, amenity proximity and normative potential. Using a spatial Big Data approach, a sample of 23761 georeferenced properties was processed by integrating web-based listings with official sources from the District (POT 555, IDECA, EPV 2024 and ICSU). To evaluate the index validity in use, a reference hedonic model was estimated; in its best functional specification it reached an explanatory power of $R^2 = 0.83$, and the incorporation of the IUG into the aggregate model lifted the adjusted R^2 by a margin that proved statistically significant under bootstrap and cross-validation. Rather than seeking predictive redundancy, the IUG was designed to be deliberately orthogonal to both price ($\rho = -0.181$) and socioeconomic stratum (AUC $= 0.483$), an independence confirmed through a protocol of eleven statistical tests including commonality analysis, spatial cross-validation by locality, Monte Carlo weight perturbation and bootstrap resampling. This statistical independence made it possible to cross the index with the residuals of the hedonic model in order to detect market inefficiencies (*mispricing*) through a 2×2 decision matrix that classifies every property into one of four actionable quadrants: buy, hold, avoid or sell. The analysis revealed that 30.7 % of the offered properties fall into the optimal investment quadrant, with prices on average 25 % lower than comparable but overvalued properties, while an additional 24.7 % concentrate in the divestment quadrant. Finally, the study is made operational through an interactive platform equipped with a conversational assistant based on a RAG architecture, democratizing data-driven real estate decision-making.

Keywords: Urban Global Index; composite synthetic indicator; hedonic model; geospatial analysis; semantic search; urban quality; Retrieval-Augmented Generation (RAG); real estate market; Bogotá.

Índice general

1. Introducción **11**

1.1. Contexto y Motivación 11

1.2. Planteamiento del IUG 12

1.3. Objetivos de la Investigación 13

1.3.1. Objetivo General 13

1.3.2. Objetivos Específicos 13

1.4. Pregunta de Investigación y Postura Metodológica 14

1.5. Hipótesis y Alcance del Proyecto 14

1.6. Contribución y Alcance Metodológico 15

2. Marco Teórico **17**

2.1. Fundamentos Econométricos 17

2.1.1. Lo que ve el mercado: Regresión Lineal Múltiple 17

2.1.2. Teoría de Precios Hedónicos 18

2.2. Enfoques de Valoración y Desarrollo Urbano 18

2.2.1. Valoración Masiva y Modelos Poblacionales 18

2.2.2. Estándares IAAO para la calidad de los avalúos masivos 18

2.2.3. El Método Residual Dinámico y su Relación con el IUG 19

2.3. Dimensiones del Bienestar Urbano 20

2.3.1. Espacio Público y Seguridad 20

2.3.2. Econometría Espacial y Machine Learning 20

2.4. Marco Conceptual y Definiciones Operativas 21

2.4.1. Definiciones Básicas del Estudio 21

2.4.2. Construcción y Formulación Matemática del IUG 21

2.4.3. Conceptos de Inteligencia Artificial 23

2.4.4. Protocolo Teórico para el Tratamiento de Ruido Espacial (DBSCAN) 24
2.4.5. Criterios de Interpretación Econométrica 24

3. Metodología **27**
3.1. Enfoque de Investigación: Big Data Urbano 27
3.2. Fuentes de Información Oficiales 28
3.3. Arquitectura del Pipeline de Datos 30
3.3.1. Metodologías, tecnologías y herramientas 31
3.3.2. Protocolo de Validación PCA 33
3.4. Consideraciones Éticas 33

4. Resultados **35**
4.1. Descripción de la Muestra 35
4.2. Subíndices del IUG 36
4.3. Índice Urbano Global (IUG) 37
4.3.1. Contraste Extremo: Mejor versus Peor Entorno 40
4.3.2. Oportunidades de Inversión 41
4.3.3. Despliegue Espacial e Inmueble Individual 41
4.4. Modelo Hedónico de Precios 42
4.5. Acceso al Sistema 47

5. Validez de Uso del IUG **49**
5.1. De la predicción a la decisión: cambio de marco 50
5.2. Bondad de ajuste y estructura del indicador 51
5.3. Robustez estadística y especificación funcional 53
5.4. Validez convergente y lectura del mercado 55
5.5. Análisis de mispricing y matriz de decisión 56
5.6. Tabla consolidada y veredicto 57

6. Conclusiones y Discusión Final: El IUG frente al Modelo Hedónico **61**

A. Arquitectura Técnica e Infraestructura **63**
A.1. Arquitectura del pipeline de datos 64
A.2. Modelo de datos y volumetría 65
A.3. Entorno computacional reproducible 67

B. Plataforma Web y Asistente Conversacional **69**

B.1. Frontend: mapa coroplético y panel de indicadores 69

B.2. Asistente conversacional (RAG) 69

B.3. Generación de cotizaciones (ACM) 71

Referencias consultadas **75**

CAPÍTULO 1

Introducción

1.1. Contexto y Motivación

El mercado de finca raíz ha constituido tradicionalmente un vehículo fundamental de ahorro e inversión en Colombia, y específicamente Bogotá se ha visto inundada por este manejo de los inmuebles. Históricamente, agentes económicos de diversa escala, desde pequeños propietarios hasta grandes fondos de inversión, han diversificado sus portafolios mediante activos inmobiliarios, bajo la premisa de que estos ofrecen estabilidad temporal, baja volatilidad y resiliencia frente a las fluctuaciones macroeconómicas.

No obstante, la valoración residencial se plantea como un desafío técnico y económico. Pues en si, la vivienda es un activo inherentemente heterogéneo y con alta especificidad espacial; su valor de mercado no solo responde a las características físicas de la estructura, sino que es función de las preferencias de los avaluadores, las condiciones del entorno inmediato y una serie de atributos urbanos (seguridad, accesibilidad, normativa) que frecuentemente resultan subrepresentados en las metodologías de valoración convencionales y hasta subjetivos. En la literatura especializada, tomando por ejemplo el enfoque hedónico que conceptualiza el valor del bien como la suma de los precios implícitos de sus características, se ha establecido como el vehículo central para explicar la formación de precios (Rosen, 1974).

Finalmente, reconociendo que la interpretación de estos datos urbanos puede ser compleja por su alta dimensionalidad el proyecto contempla una fase de despliegue

tecnológico mediante metodologías de Inteligencia Artificial Generativa. En la que se aprovecha la capacidad de modelos de lenguaje avanzados y la riqueza de los datos no estructurados de los inmuebles para habilitar un asistente conversacional, permitiendo consultar un indicador que represente información objetiva y encontrar oportunidades de inversión mediante lenguaje natural.

1.2. Planteamiento del IUG

En este contexto, el presente trabajo propone el desarrollo del Índice Urbano Global (IUG) como una herramienta de análisis estadístico poblacional para Bogotá. La base empírica del estudio compila un extenso dataset de anuncios de vivienda publicados en portales de alta penetración en el mercado. Si bien esta muestra no alcanza el universo transaccional total, captura una porción estadísticamente significativa de la oferta efectiva, permitiendo una trazabilidad espacial detallada a nivel de Barrio y Localidad.

El IUG contribuye no solo a explicar la varianza en los precios, sino a medir la calidad integral del activo inmobiliario de manera independiente al valor monetario. El índice sintetiza cinco dimensiones críticas:

- **Componente Hedónico (Calidad Intrínseca):** Evalúa las especificaciones propias del inmueble, considerando tanto su dimensión espacial (área construida, distribución) como sus atributos físicos y de confort.

- **Seguridad objetiva y entorno:** Medición basada en datos históricos de incidencia delictiva (GeoJSON) y percepción ciudadana.

- **Accesibilidad y movilidad:** Oferta de transporte público (SITP/TransMilenio) y conectividad vial.

- **Proximidad a puntos de interés (Dotacionales):** Cercanía a servicios de salud, educación, abastecimiento y espacio público.

- **Potencial normativo:** Capacidad de edificabilidad y lineamientos de planeación urbana (POT).

1.3. Objetivos de la Investigación

1.3.1. Objetivo General

Diseñar, implementar y validar el **IUG** como una herramienta poblacional, abierta y reproducible que sintetice cinco dimensiones del bienestar urbano de Bogotá en una métrica única. Este indicador operará como un sistema de soporte a la decisión inmobiliaria, calculado de forma estrictamente independiente al precio de mercado.

1.3.2. Objetivos Específicos

1. **Operacionalizar** las dimensiones de accesibilidad, seguridad, calidad construida, dotación y potencial normativo en subíndices comparables, utilizando fuentes oficiales abiertas (DANE, IDECA, SDP, POT 555, EPV 2024, ICSU).

2. **Construir un pipeline de datos reproducible** que integre la extracción automatizada de oferta inmobiliaria, asignaciones espaciales a nivel de bases de datos, control de calidad mediante algoritmos de agrupación y persistencia versionada, garantizando la trazabilidad desde el anuncio hasta el indicador final.

3. **Estimar un modelo hedónico de referencia** (OLS, Ridge, Lasso, Elastic Net) cuyo propósito exclusivo es servir como *base line* descriptivo para contrastar la información incremental que aporta el IUG.

4. **Validar estadísticamente el indicador** mediante un protocolo de once pruebas que evalúe su bondad de ajuste, robustez, validez convergente y de uso.

5. **Cuantificar la incertidumbre** del índice mediante descomposición de varianza (índices de Sobol) y análisis de sensibilidad ante perturbación Monte Carlo de sus pesos.

6. **Habilitar el acceso operativo** mediante un despliegue web público y un asistente conversacional (RAG) que democratice la consulta del indicador sin requerir conocimientos técnicos en geoestadística.

1.4. Pregunta de Investigación y Postura Metodológica

Para materializar estos objetivos, es fundamental resolver la siguiente cuestión central: *¿Es posible construir, a partir de fuentes abiertas y replicables, un indicador sintético de calidad urbana que sea estadísticamente válido, robusto y operativamente útil para identificar discrepancias entre el precio observado y los fundamentales urbanos del territorio?*

Para responderla, esta tesis asume una postura metodológica deliberadamente *descriptiva-decisional*, y no predictiva. Un señalamiento natural sobre cualquier métrica urbana es cuestionar qué tan bien predice el precio. Sin embargo, el IUG **no se propone inferir, estimar ni predecir el precio de mercado** de un inmueble. Ya existen familias maduras de Modelos de Valoración Automática (AVM) que cumplen ese rol con métricas establecidas por la *International Association of Assessing Officers* (IAAO, 2013). Estos estándares (COD, PRD y PRB) se desarrollan con detalle en la Sección 2.2.2 del Marco Teórico, y se retoman como diagnóstico del modelo hedónico de referencia en el Capítulo 4.

La propuesta del IUG es funcionar como una **herramienta-perspectiva**: un instrumento analítico que condensa cinco metodologías heterogéneas en una escala interpretable y ortogonal al precio. Su virtud no reside en la correlación con el valor comercial, sino en su capacidad de revelar la brecha existente entre lo que el mercado paga y la calidad objetiva del entorno.

1.5. Hipótesis y Alcance del Proyecto

Este enfoque se sustenta en una **hipótesis central**: existe información estructural sobre la calidad urbana que el precio no captura íntegramente. A nivel operativo, se plantea que la ortogonalidad entre el IUG y el precio (con una correlación débil esperada de $|\rho| < 0.30$) no es un defecto del modelo, sino la condición necesaria para que este aporte información nueva sobre el *benchmark* hedónico.

Para delimitar con precisión el fin y alcance de este proyecto, es importante recalcar que esta tesis:

- No sustituye la valoración profesional caso a caso ni estima precios puntuales para inmuebles individuales.

- No incorpora costos de obra, tasas de descuento ni evalúa estructuras financieras de proyectos (Método Residual Dinámico).
- No pretende fungir como instrumento tributario ni base catastral, ofreciendo un complemento analítico en lugar de una sustitución regulatoria.
- No pretende alcanzar eficiencia predictiva sobre el precio, ya que la divergencia con las métricas tradicionales es el núcleo que permite identificar discrepancias de valor.

1.6. Contribución y Alcance Metodológico

La idea central de la investigación es comparar el precio observado con un conjunto de fundamentales urbanos para reducir la subjetividad inherente a las valoraciones zonales. El alcance metodológico integra cuatro componentes: un *benchmark* hedónico para establecer la línea base, la construcción del índice sintético multidimensional, un protocolo de validación estadística y un despliegue tecnológico final.

La relación entre las dimensiones de calidad y el mercado se materializa mediante el Ratio de Oportunidad, un instrumento diseñado para detectar *mispricing* relativo:

$$\text{Ratio} = \frac{IUG}{\text{Precio Estandarizado}} \tag{1.1}$$

Esta métrica ofrece valor directo a inversionistas y al sector financiero, permitiendo la identificación *data-driven* de oportunidades (alto valor urbanístico y precios contenidos) o la emisión de alertas (baja calidad y precios elevados). Adicionalmente, subíndices específicos como el de seguridad (I_{SEG}) tienen aplicabilidad directa en el ámbito actuarial para modelos de frecuencia, severidad y tarificación geoespacial.

Reproducibilidad y código abierto. El código fuente, la documentación técnica detallada, las instrucciones de despliegue local y las capturas representativas del sistema están disponibles públicamente en el repositorio del proyecto:

`https://github.com/neylinsomne/indice-urbano-global-bogota`

La plataforma desplegada es accesible en `https://inmu.vara-alta.lat`. El detalle técnico completo se documenta en el Anexo A.

CAPÍTULO 2

Marco Teórico

2.1. Fundamentos Econométricos

2.1.1. Lo que ve el mercado: Regresión Lineal Múltiple

El punto de partida para modelar el precio es la especificación lineal clásica. Sea Y la variable respuesta (precio o log-precio) y $X_1, \ldots, X_p$ los predictores (atributos). En notación matricial, el modelo se expresa como:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

Donde:

- $\mathbf{X} \in \mathbb{R}^{n \times (p+1)}$ es la matriz de diseño (incluyendo el intercepto).
- $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^\top$ es el vector de coeficientes a estimar.
- $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$ es el término de error, bajo supuestos de independencia y homocedasticidad.

El estimador de Mínimos Cuadrados Ordinarios (MCO) está dado por $\hat{\boldsymbol{\beta}} = (\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top\mathbf{Y}$, siendo este el Mejor Estimador Lineal Insesgado (BLUE) bajo las condiciones de Gauss-Markov. En la práctica inmobiliaria, es crítico evaluar la linealidad, la normalidad de los residuos y, especialmente, la multicolinealidad entre atributos estructurales.

2.1.2. Teoría de Precios Hedónicos

La complejidad de la naturaleza de los activos inmobiliarios hace que su valoración sea un desafío particular. Para abordar esta heterogeneidad, la Teoría de Precios Hedónicos, formalizada por Rosen (1974), plantea que un bien no se valora como un elemento indivisible, sino por el conjunto de utilidades que ofrece. De este modo, el modelo descompone el precio total en la suma de los valores implícitos de sus dimensiones: desde las características estructurales propias del inmueble (como el área, la edad y los acabados), pasando por las dinámicas del vecindario (estrato y demografía), hasta llegar a su localización, la cual define su nivel de accesibilidad y cercanía a los centros de actividad económica (CBD).

Mientras que la estimación por MCO captura los efectos promedio, la **Regresión Cuantílica** permite revelar la heterogeneidad de estos efectos a lo largo de la distribución condicional del precio (ej. en los cuantiles q_{10}, q_{50}, q_{90}), reconociendo que el mercado de lujo puede valorar los atributos de forma distinta al mercado de vivienda de interés social.

2.2. Enfoques de Valoración y Desarrollo Urbano

2.2.1. Valoración Masiva y Modelos Poblacionales

En diversas jurisdicciones, la gestión tributaria y catastral ha migrado hacia esquemas de *Valoración Masiva Automatizada* (AVM). Estos sistemas buscan consistencia y eficiencia mediante modelos econométricos calibrados con información transaccional y catastral. Este trabajo se alinea con este espíritu comparativo al estimar un modelo supervisado para caracterizar "lo que ve el mercado", pero innova al proponer el IUG como una métrica independiente del precio para contrastar el valor observado con los fundamentales urbanos.

2.2.2. Estándares IAAO para la calidad de los avalúos masivos

La *International Association of Assessing Officers* (IAAO) publica los estándares internacionales más extendidos para evaluar la calidad estadística de modelos de valoración

masiva (IAAO, 2013). Estos estándares se materializan en tres métricas que el Capítulo 4 retoma al diagnosticar el modelo hedónico:

COD (*Coefficient of Dispersion*): mide la *uniformidad* de las estimaciones. Se calcula como el promedio de las desviaciones absolutas del cociente (precio estimado / precio observado) respecto a su mediana, expresado en porcentaje. Un COD bajo indica que el modelo trata de forma homogénea a inmuebles similares. La IAAO considera **aceptable un COD entre 5 % y 15 %** para vivienda residencial; valores superiores reflejan dispersión inherente al mercado que ningún modelo puede comprimir con datos de oferta.

PRD (*Price-Related Differential*): mide la *equidad vertical* (*progresividad / regresividad*). Se obtiene como el cociente entre la media aritmética y la media ponderada por valor de los ratios precio estimado/observado. Un **PRD en el rango** $[0.98,\ 1.03]$ indica neutralidad; valores superiores a 1.03 sugieren *regresividad* (el modelo subvalora propiedades caras frente a baratas), mientras que los inferiores a 0.98 sugieren *progresividad*.

PRB (*Price-Related Bias*): complementa al PRD con una medida más robusta de la misma equidad vertical, estimada como la pendiente de una regresión auxiliar entre el ratio precio/valor y el logaritmo del valor. La IAAO considera aceptables **valores entre** -0.05 **y** $+0.05$; pendientes cuya magnitud sea mayor confirman un sesgo sistemático respecto al nivel de precio.

Estos tres estadísticos se complementan con el *ratio mediano* (precio estimado / precio observado), que debe situarse en torno a 1.00 para garantizar la ausencia de sesgo global. En el presente trabajo se reportan los cuatro valores como diagnóstico del modelo hedónico (Cuadro 4.13) y no como criterio de aceptación del IUG, dado que este último no se propone como AVM.

2.2.3. El Método Residual Dinámico y su Relación con el IUG

En el ámbito del desarrollo inmobiliario, la valoración de suelo o proyectos en curso suele realizarse mediante el *Método Residual Dinámico*. Este enfoque calcula el valor actual del suelo partiendo del valor hipotético del proyecto terminado y descontando

los costos de construcción, honorarios, financieros y la utilidad del promotor en el tiempo (Tasadorfiable, 2025). Es especialmente útil para activos que aún no generan rentas o suelos urbanizables.

Nota 1 (**Relación con este estudio**). El enfoque residual es idóneo cuando el *potencial edificatorio* es el driver principal del valor. En esta tesis, dicha dimensión se captura mediante el subíndice I_{PNU} (Potencial Normativo–Urbanístico), que resume la capacidad edificatoria (índice de construcción) y los usos permitidos por el POT.
Sin embargo, el IUG no internaliza explícitamente costos de obra o tasas de descuento financiero; su función es complementaria. Mientras el método residual determina la viabilidad financiera de un proyecto específico, el IUG y el ratio IUG/Precio ofrecen un *baseline* espacial poblacional para detectar si el mercado está pagando por el suelo un precio acorde a su capacidad normativa y calidad urbana, independientemente de la estructura de costos de un desarrollador particular.

2.3. Dimensiones del Bienestar Urbano

2.3.1. Espacio Público y Seguridad

El bienestar urbano trasciende la propiedad privada. El espacio público (zonas verdes, parques) y la seguridad (tanto objetiva, medida en tasas de criminalidad, como subjetiva/percibida) son determinantes en la disposición a pagar de los hogares. La literatura sugiere que la inseguridad actúa como un "impuesto implícito" sobre el valor de la vivienda, mientras que la accesibilidad peatonal y el transporte masivo generan plusvalías por localización.

2.3.2. Econometría Espacial y Machine Learning

Dado que "todo está relacionado con todo, pero las cosas cercanas están más relacionadas que las distantes"(Primera Ley de la Geografía, Tobler), los modelos inmobiliarios deben gestionar la dependencia espacial.

- **Diagnóstico:** Uso del Índice de Moran para detectar autocorrelación en los residuos.

- **Exploración:** Técnicas de reducción de dimensionalidad como **PCA** (Análisis de Componentes Principales) para reducir la dimensionalidad de atributos correlacionados, y algoritmos de agrupamiento como **DBSCAN** para identificar "clústeres" espaciales de precios o anomalías que no siguen la lógica del mercado general.

2.4. Marco Conceptual y Definiciones Operativas

A diferencia del marco teórico, que expone los fundamentos científicos generales, esta sección operacionaliza dichas teorías en variables específicas, estableciendo el sistema de ecuaciones del Índice Urbano Global (IUG) y los protocolos de manejo de datos.

2.4.1. Definiciones Básicas del Estudio

Inmueble (Unidad de Análisis): Anuncio georreferenciado de vivienda (Propiedad Horizontal o Casa) en Bogotá. El precio observado se estandariza como precio por metro cuadrado ($\tilde{P}_i$), ajustado por métricas de control.

Barrio / Localidad: Unidades administrativas del Distrito Capital usadas como contenedores espaciales para la agregación de métricas zonales y validación geométrica.

Modelo Hedónico: Especificación econométrica supervisada que explica el logaritmo del precio en función de los atributos. Representa *lo que ve el mercado*.

Índice Urbano Global (IUG): Indicador sintético adimensional (escala 0-5), *independiente del precio*, que cuantifica la calidad integral del entorno urbano.

2.4.2. Construcción y Formulación Matemática del IUG

El modelo conceptual del índice se define como la combinación lineal ponderada de cinco dimensiones críticas del bienestar urbano. Para cada inmueble i, la formulación matemática se expresa como:

$$IUG_i = w_1 I_{HED} + w_2 I_{ACC} + w_3 I_{SEG} + w_4 I_{DOT} + w_5 I_{PNU} \tag{2.2}$$

Donde $\sum w_k = 1$. Cada uno de estos componentes (I_k) se calcula a partir de variables operativas específicas, detalladas en la Tabla 2.1.

Cuadro 2.1: Definición operativa y dirección de los subíndices del IUG.

Símbolo	**Dimensión**	**Definición Operativa**	**Signo (S)**
I_{HED}	Calidad Intrínseca	**Score Hedónico:** Resultado de un PCA sobre variables estructurales (área, baños, garajes). Representa el confort privado.	+
I_{ACC}	Accesibilidad	**Modelo Gravitacional:** Sumatoria de oportunidades ponderadas por el inverso de la distancia ($d_{ij}^{-\beta}$) a estaciones de TransMilenio y SITP.	+
I_{SEG}	Seguridad	**Indicador Tridimensional:** Combina tres capas: (1) de proximidad local (proximidad a CAI y sectores priorizados, peso 0.40), (2) criminalidad objetiva ICSU con 7 tipos de delito normalizados por km^2 (peso 0.30), y (3) percepción subjetiva derivada de la EPV 2024 (peso 0.30).	$-$ (Inv)
I_{DOT}	Dotacional	**Proximidad Acumulada:** Disponibilidad espacial de equipamientos (salud, educación, cultura) en radios de caminabilidad (400m - 800m).	+
I_{PNU}	Normativa	**Potencial de Desarrollo:** Scoring basado en el POT 555, evaluando la edificabilidad permitida (altura) y el tratamiento urbanístico.	+

Se seleccionaron estas métricas bajo criterios de disponibilidad de datos abiertos y literatura previa. Específicamente, para la dimensión de **Seguridad** (I_{SEG}), el diseño original contemplaba un enfoque de densidad inversa basado exclusivamente en datos objetivos de criminalidad. Durante la fase de desarrollo, la auditoría metodológica reveló que la correlación entre criminalidad objetiva y percepción ciudadana es débil ($r = 0.3174$), y que el 44 % de las localidades exhibe la denominada "paradoja de inseguridad" (baja criminalidad reportada pero alta percepción de riesgo). Esto motivó la reformulación hacia un indicador tridimensional:

$$I'_{SEG} = \underbrace{0.40 \cdot S_{micro}}_{\text{CAI + sectores}} + \underbrace{0.30 \cdot D_{obj}}_{\text{criminalidad ICSU}} + \underbrace{0.30 \cdot D_{subj}}_{\text{percepcion EPV}} \tag{2.3}$$

donde D_{obj} proviene del Indicador Compuesto de Seguridad Urbana (ICSU), que analiza 7 tipos de delito normalizados por km^2, y D_{subj} integra 8 variables de la Encuesta de Percepción y Victimización (EPV 2024, Cámara de Comercio de Bogotá, $n = 19\,354$). Los pesos son configurables mediante perfiles de usuario (equilibrado, inversión, residencial), lo que permite que distintos perfiles de decisión prioricen la dimensión más relevante para su caso de uso. El componente de proximidad local (S_{micro}) se preserva como eje articulador del indicador.

Para el componente **Dotacional** (I_{DOT}), los radios de 400m y 800m no son arbitrarios; corresponden a los estándares de caminabilidad urbana (5 y 10 minutos a pie) que determinan la cobertura efectiva de servicios para un residente sin vehículo particular.

Normalización y Escala Común Dado que las dimensiones descritas anteriormente poseen unidades heterogéneas (metros, densidades, conteos), es necesario llevarlas a una escala uniforme. Se adopta una normalización Min-Max escalada al rango $[0, 5]$ con el objetivo de facilitar la interpretación del usuario final:

$$S(X_i) = 5 \times \frac{\text{rank}(X_i) - 1}{N - 1} \tag{2.4}$$

A diferencia de una estandarización Z-score (que produce valores negativos difíciles de comunicar a un público no experto), esta escala permite una lectura intuitiva donde 5 representa la máxima dotación teórica observada en la muestra. Finalmente, esta métrica permite calcular el **Ratio de Oportunidad**:

$$\text{Ratio}_i = \frac{IUG_i}{\tilde{P}_i} \tag{2.5}$$

Donde un Ratio > 1 sugiere una subvaloración relativa, indicando una potencial oportunidad de inversión.

2.4.3. Conceptos de Inteligencia Artificial

Para sustentar el módulo de despliegue tecnológico (Chatbot) descrito en la metodología, se definen los siguientes conceptos fundamentales:

Embeddings (Vectorización): Representación matemática de texto no estructurado en vectores numéricos densos, donde la cercanía geométrica implica similitud

semántica.

RAG (Retrieval-Augmented Generation): Arquitectura que conecta modelos de lenguaje (LLMs) con bases de datos vectoriales privadas para responder consultas sobre el índice con veracidad.

2.4.4. Protocolo Teórico para el Tratamiento de Ruido Espacial (DBSCAN)

El tratamiento del ruido inherente a los datos de portales inmobiliarios exige una delimitación conceptual clara para la detección de valores atípicos. Para este fin, se fundamenta el uso del algoritmo de clustering espacial basado en densidad (DBSCAN). Teóricamente, las observaciones que este algoritmo aísla como ruido (típicamente etiquetadas por el sistema como `-1`) representan posibles errores de geocodificación o anomalías extremas que no siguen la lógica del mercado general.

Para preservar la integridad del análisis, se establece un principio de separación dual: los modelos hedónicos deben estimarse exclusivamente sobre la muestra limpia para prevenir sesgos estructurales en los estimadores paramétricos; en contraste, el cálculo de la calidad urbana sí debe abarcar la totalidad del territorio, manteniendo los atípicos visibles bajo una simbología diferenciada para alertar al decisor sobre estas anomalías sin distorsionar la econometría base.

2.4.5. Criterios de Interpretación Econométrica

La evaluación de los modelos de precios trasciende la simple bondad de ajuste global. Conceptualmente, la lectura de los resultados econométricos se estructura sobre tres dimensiones analíticas: En primer lugar, se evalúa la **señal y magnitud** económica de los estimadores obtenidos; en segundo lugar, se verifica la **precisión estadística** mediante el reporte de intervalos de confianza y errores estándar robustos a la heterocedasticidad (HC); y finalmente, se explora la **heterogeneidad** del mercado, reconociendo que los atributos urbanos pueden valorarse de forma distinta en los extremos de la distribución de precios (por ejemplo, contrastando los deciles q_{10} y q_{90}).

Validación Cruzada Espacial y la Primera Ley de la Geografía Un desafío teórico insoslayable en la modelación inmobiliaria es la autocorrelación espacial. De acuerdo con la Primera Ley de la Geografía postulada por Tobler, los inmuebles cercanos tienden a exhibir precios similares por compartir dinámicas de entorno. En este contexto, una validación cruzada estándar (K-Fold aleatorio) resulta teóricamente deficiente, ya que induce *data leakage* al separar observaciones de un mismo edificio o cuadra en pliegues distintos, simulando un poder predictivo artificial.

Para mitigar este riesgo, se conceptualiza un esquema de validación cruzada espacial mediante agrupamiento (**GroupKFold por localidad**), donde el bloque geográfico completo se mantiene indivisible. La magnitud de esta fuga de información se cuantifica mediante la siguiente razón de penalización (*leakage ratio*):

$$\text{Leakage} = \frac{\text{RMSE}_{\text{espacial}} - \text{RMSE}_{\text{estandar}}}{\text{RMSE}_{\text{estandar}}} \tag{2.6}$$

Si este incremento supera un umbral de tolerancia del 15 %, se asume estadísticamente que el modelo base está memorizando efectos locales de barrio en lugar de aprender patrones estructurales que sean verdaderamente generalizables a nuevas observaciones.

Intervalos de Confianza con t-Student en el ACM El módulo de Análisis Comparativo de Mercado (ACM) requiere estimar intervalos de confianza a partir de conjuntos pequeños de comparables ($n = 2$–10). Con muestras de este tamaño, la distribución normal subestima la incertidumbre del estimador, produciendo intervalos artificialmente estrechos. Por tanto, se adopta la distribución t-Student con $n - 1$ grados de libertad y desviación típica *muestral* ($\div(n-1)$) para el cálculo de los intervalos al 90 %. A modo de referencia, para $n = 3$ el valor crítico $t_{2,\,0.95} = 2.920$ produce intervalos un 77 % más anchos que los obtenidos con $z_{0.95} = 1.645$, reflejando con mayor fidelidad la incertidumbre real de la estimación. La asimetría del intervalo solo se reporta cuando $n \geq 8$.

CAPÍTULO 3

Metodología

El presente capítulo detalla la estrategia de ejecución técnica adoptada para la construcción del Índice Urbano Global (IUG). A diferencia del marco conceptual, que define las variables teóricas, esta sección describe la **arquitectura de datos**, el **flujo de procesamiento** y las **herramientas computacionales** que permiten transformar los datos brutos en métricas de valor, garantizando la reproducibilidad, escalabilidad y auditoría del estudio.

3.1. Enfoque de Investigación: Big Data Urbano

El estudio se enmarca dentro del paradigma de **Big Data** y sigue la metodología estándar **CRISP-DM** (Cross-Industry Standard Process for Data Mining). Según la definición canónica de las "3 Vs"(Laney, 2001), el proyecto cumple con los criterios taxonómicos de esta disciplina:

- **Volumen:** Gestión de cartografía catastral masiva (millones de vértices geométricos) y bases históricas de seguridad con alta densidad de registros.
- **Variedad:** Fusión de formatos heterogéneos, integrando archivos espaciales (`SHP`, `GeoJSON`), registros tabulares históricos (`CSV` de delitos) y datos no estructurados (`HTML` proveniente de Web Scraping).

- **Veracidad:** Validación cruzada de la oferta de mercado contra fuentes oficiales del Distrito (IDECA, SDP) para mitigar el ruido inherente a la información web.

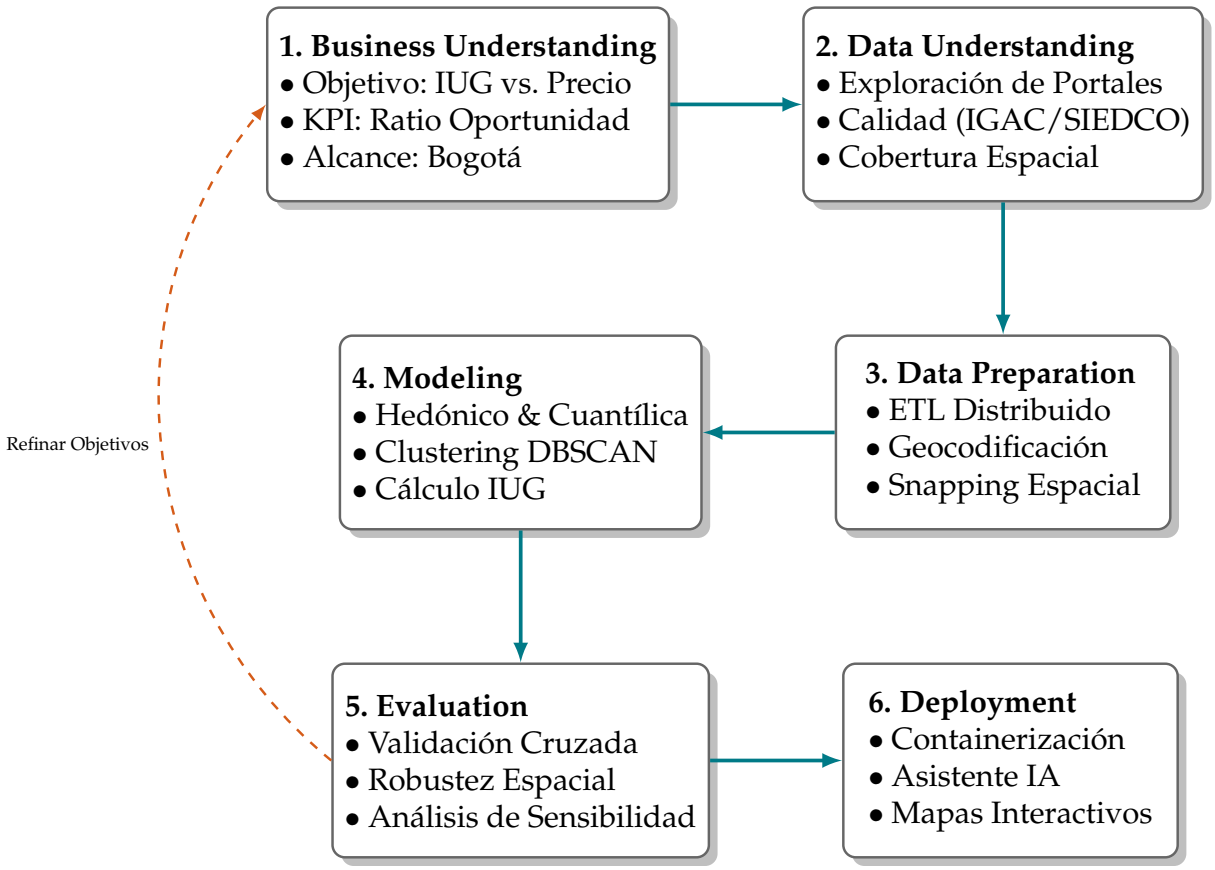


Figura 3.1: Ciclo de vida CRISP-DM adaptado al proyecto IUG: Flujo iterativo desde la definición del negocio hasta el despliegue del asistente virtual.

3.2. Fuentes de Información Oficiales

La construcción del IUG se fundamenta en la integración de datos abiertos gubernamentales y fuentes colaborativas, asegurando la calidad y legalidad de la información base. La Tabla 3.1 detalla el inventario de datasets utilizados.

La muestra consolidada de 23761 inmuebles bogotanos, ingeridos desde 45438 registros multi-ciudad totales (la base incluye 21671 registros de Medellín, Cali y Barranquilla que quedan fuera del alcance espacial del estudio), supera ampliamente los umbrales mínimos para la convergencia de los estimadores en los modelos de regresión hedónica planteados, garantizando, por el Teorema del Límite Central, que los resultados posean plena validez econométrica.

Cuadro 3.1: Inventario detallado de fuentes de datos espaciales y alfanuméricos.

Dimensión	Dataset / Capa	Formato	Fuente
1. Normativa y Administrativo (SDP / IDECA)			
División	Localidades, UPL y Sectores Catastrales	GeoJSON	Datos Abiertos Bogotá
Normativa (Piso)	Tratamientos Urbanísticos (POT 555)	SHP	Secretaría Planeación
Edificabilidad	Rangos de edificabilidad y Alturas	SHP	Datos Abiertos Bogotá
Usos del Suelo	Áreas de Actividad (Residencial/Comercial)	GDB	IDECA
2. Seguridad y Convivencia (Secretaría de Seguridad / Policía)			
Zonas	Cuadrantes de Policía y Sectores Priorizados	GeoJSON	Datos Abiertos Bogotá
Puntos	Estaciones de Policía y CAI	SHP (ZIP)	IDECA
Incidencia	Delitos de Alto Impacto (Mapas de Calor)	CSV	Sec. Seguridad
Percepción	Encuesta de Percepción y Victimización (EPV 2024)	Tabulados	Cámara de Comercio
Barrios	Barrios Legalizados	JSON	Datos Abiertos Bogotá
3. Equipamientos y Servicios (Dotacional)			
Educación	Colegios (Sedes) y Universidades	CSV / GeoJSON	Sec. Educación
Salud	IPS y Farmacéuticos (Droguerías)	SHP / GeoJSON	Datos Abiertos Bogotá
Abastecimiento	Plazas de Mercado (IPES)	Excel / GeoJSON	IPES
Comercio	Centros Comerciales y Supermercados	CSV / OSM	Kaggle / OpenStreetMap
Cultura	Red de Bibliotecas (BibloRed) y Teatros	GeoJSON	Sec. Cultura
Recreación	Parques y Canchas (IDRD)	SHP / CSV	IDRD
4. Mercado Inmobiliario			
Oferta	Anuncios de Vivienda (Venta/Arriendo)	HTML (DOM)	Web Scraping

3.3. Arquitectura del Pipeline de Datos

Para sostener un análisis poblacional sobre miles de inmuebles y decenas de capas oficiales, fue necesario diseñar un *pipeline* automatizado capaz de ingerir, limpiar, georreferenciar, integrar y servir información heterogénea de forma reproducible. La lógica del pipeline se organiza en tres etapas secuenciales que aíslan responsabilidades: **ingesta** de datos crudos, **procesamiento** y enriquecimiento espacial, y **distribución** hacia dos almacenes especializados.

La etapa de **ingesta** concentra dos flujos de origen distinto. Por un lado, los anuncios de oferta inmobiliaria se recolectan mediante *web scraping* programado contra portales de alta penetración en el mercado bogotano; por otro, las capas oficiales (POT 555, dotaciones IDECA, criminalidad por localidad, EPV 2024, paraderos de transporte público) se descargan periódicamente desde los portales de datos abiertos. La heterogeneidad de formatos resultantes (HTML, GeoJSON, SHP, CSV, JSONB) se preserva en una etapa de *staging* sobre MongoDB, lo que permite reprocesar el pipeline sin tener que volver a recolectar la fuente original.

La etapa de **procesamiento** normaliza atributos, depura registros mediante DBSCAN para detectar outliers, **geocodifica** los anuncios y los asigna espacialmente a barrios, localidades y polígonos del POT vía operadores PostGIS de tipo `ST_Within` y `ST_Distance`. La *geocodificación* es el proceso por el cual una dirección textual escrita en lenguaje natural (e.g. "Calle 100 # 15-50, Chapinero") se transforma en un par de coordenadas geográficas (latitud, longitud); en este pipeline se resuelve consultando el servicio Nominatim sobre OpenStreetMap, con un *viewbox* restringido al perímetro de Bogotá para evitar coincidencias falsas en otras ciudades. Es en este punto donde un anuncio textual, como "apartamento amplio en el norte", se convierte en un registro georreferenciado con coordenadas, barrio asignado y un perfil de proximidad a equipamientos urbanos.

Tras el procesamiento, el pipeline **distribuye** la información hacia dos almacenes con propósitos complementarios. La **rama analítica espacial**, sobre PostGIS, alimenta el cálculo de los cinco subíndices y los modelos econométricos del Capítulo 4; aquí residen las consultas SQL parametrizadas que el frontend invoca para construir el mapa coroplético y las tablas de ranking. La **rama vectorial**, sobre PgVector, transforma

la información no estructurada de los inmuebles (descripciones textuales, atributos, contexto de barrio) en representaciones densas (*embeddings*) que habilitan consultas semánticas del tipo *"apartamentos cerca a parques seguros"*, algo imposible de resolver con un filtro SQL convencional pero natural para un asistente conversacional basado en RAG.

La separación entre ambas ramas es deliberada: la analítica exige consistencia, trazabilidad y auditoría sobre cada cifra reportada; la vectorial exige flexibilidad semántica y tolerancia a la ambigüedad del lenguaje natural. Compartir el origen pero diversificar el almacén permite atender las dos demandas sin compromisos cruzados.

El detalle técnico exhaustivo (que incluye el diagrama completo del pipeline, modelo de datos por tabla, volumetría, configuración de contenedores Docker y orquestación de tareas mediante GitHub Actions) se documenta en el Anexo A; este capítulo conserva la mirada metodológica para que el lector pueda evaluar las decisiones de diseño sin perderse en su implementación.

3.3.1. Metodologías, tecnologías y herramientas

El sistema se construyó sobre un *stack* abierto y reproducible. La Tabla 3.2 sintetiza las herramientas seleccionadas y la responsabilidad concreta de cada una dentro del pipeline.

La elección del *stack* obedece a tres criterios articulados. Primero, **reproducibilidad**: cada componente es de código abierto y la pila completa puede levantarse con un único comando `docker compose up`, lo que blinda el trabajo frente a la obsolescencia de servicios propietarios. Segundo, **trazabilidad espacial**: PostGIS es el estándar *de facto* para análisis geoespacial, y su combinación con PgVector en un mismo motor PostgreSQL evita la fragmentación de datos entre múltiples bases. Tercero, **escalabilidad metodológica**: el uso de Flyway garantiza que cualquier cambio en el cálculo de indicadores quede versionado y sea auditable, condición indispensable para que un trabajo académico sea verificable por terceros.

El código fuente, las migraciones, los scripts de carga y la documentación técnica están disponibles públicamente en el repositorio del proyecto: https://github.com/neylinsomne/indice-urbano-global-bogota.

Cuadro 3.2: Stack tecnológico del proyecto.

Capa	Herramienta	Responsabilidad
Datos y persistencia		
Almacenamiento crudo	MongoDB	*Staging* de anuncios y datos abiertos heterogéneos
Analítica espacial	PostgreSQL 16 + PostGIS	Cálculo de indicadores, modelos hedónicos, agregaciones
Búsqueda semántica	PgVector	Embeddings densos para asistente RAG
Versionado de esquema	Flyway	Migraciones idempotentes y auditables
Cache y cola	Redis	Sesiones, cache de consultas, jobs asíncronos
Procesamiento y modelos		
Web scraping	Scrapy + Playwright	Extracción de anuncios con renderizado JavaScript
Modelado estadístico	scikit-learn, statsmodels	OLS, Ridge, Lasso, Elastic Net, PCA, DBSCAN
Cómputo espacial	GeoPandas, Shapely	Geocodificación e intersecciones
API y frontend		
Backend	FastAPI (Python 3.12)	API REST + endpoints analíticos del Capítulo 5
Frontend	React + Vite	Interfaz interactiva
Cartografía	Leaflet + react-leaflet	Mapa coroplético multinivel
LLM y RAG	Gemini	Asistente conversacional sobre PgVector
Despliegue y operación		
Contenedores	Docker Compose	Orquestación local y producción
Túnel público	Cloudflare Tunnel	Exposición segura sin abrir puertos
CI/CD y automatización	GitHub Actions (runner self-hosted)	Scraping mensual y recálculo de indicadores
Monitoreo	Prometheus + Grafana	Métricas de pipeline y disponibilidad

3.3.2. Protocolo de Validación PCA

Dado que varios subíndices del IUG emplean Análisis de Componentes Principales para la extracción de pesos, tanto en el score hedónico (I_{HED}) como en la ponderación de modos de transporte (I_{ACC}), resulta indispensable verificar que la estructura de correlación de los datos justifique la reducción factorial antes de ejecutarla. Se implementa un protocolo de validación obligatorio previo a cualquier PCA, compuesto por cuatro pasos: (1) el test KMO (Kaiser-Meyer-Olkin), que verifica que las correlaciones parciales son lo suficientemente pequeñas, con un umbral mínimo de KMO ≥ 0.50; (2) el test de esfericidad de Bartlett, que verifica que la matriz de correlación difiere significativamente de la identidad ($p < 0.05$); (3) el criterio de Kaiser, que retiene únicamente componentes con autovalor $\lambda > 1$; y (4) un bootstrap de estabilidad con 200 remuestreos para calcular intervalos de confianza de los *loadings*.

El mínimo de observaciones para ejecutar un PCA se estableció en $n = 40$, garantizando una razón $n/p \geq 10$ con 4 variables de transporte. Si el PCA no supera la validación en cualquiera de los primeros dos tests, se activa una jerarquía de *fallback* automática: PCA robusto (RobustScaler) $\rightarrow$ pesos proporcionales a la varianza $\rightarrow$ promedio simple. Este mecanismo asegura que ningún subíndice quede sin calcular por insuficiencia muestral, pero que tampoco se produzcan componentes dominados por ruido aleatorio cuando la estructura de datos es inadecuada para la factorización.

3.4. Consideraciones Éticas

El proyecto se adhiere estrictamente a la Ley 1581 de 2012 sobre protección de datos. Los registros de la base de datos en ningún momento almacenan información privada ni personal de los propietarios o anunciantes. Se utilizan exclusivamente datasets públicos o recolectados bajo políticas de uso justo con fines académicos y de investigación.

CAPÍTULO 4

Resultados

Este capítulo presenta los hallazgos del estudio organizados en cinco ejes: la muestra final, los subíndices y el IUG desplegado espacialmente, los modelos hedónicos de precios, el asistente conversacional y, finalmente, la discusión comparativa entre lo que el mercado refleja a través de la regresión y lo que el IUG mide como calidad urbana.

4.1. Descripción de la Muestra

Fecha de corte y vigencia. Los datos de mercado corresponden al corte mensual del pipeline automatizado vigente al cierre de mayo de 2026, alimentado por el ciclo de *scraping* programado el día 15 de cada mes. Las capas oficiales (POT 555, criminalidad por localidad, EPV 2024, dotaciones IDECA) corresponden a sus versiones públicas más recientes a esa misma fecha. El estudio admite replicación con cortes posteriores sin cambio metodológico: el resultado de cualquier mes futuro se obtiene re-ejecutando el *pipeline* sobre el dataset actualizado.

Tras la migración desde MongoDB Atlas y la asignación espacial mediante PostGIS, el dataset final comprende **17547 inmuebles** de vivienda en Bogotá, distribuidos en 20 localidades y 1721 barrios legalizados. La muestra mantiene la composición tipológica característica del mercado bogotano: el 51.6 % corresponde a apartamentos y el 48.4 % restante a casas. El Cuadro 4.1 resume los descriptivos por tipología.

El DBSCAN identificó 2510 outliers (14.3 % del corpus total georreferenciado),

Cuadro 4.1: Distribución de la muestra de vivienda bogotana.

Tipo	n	**Precio/m^2 mediano**	**Área med.**	**Estrato**	**%**
Apartamento	9052	\$5 200 000	85 m^2	4	51.6 %
Casa	8495	\$3 900 000	180 m^2	3	48.4 %
Total	**17547**				**100 %**

agrupando posibles errores de geocodificación y precios atípicos. Estos se excluyeron de los modelos hedónicos para evitar sesgos en los estimadores, pero se conservaron en el cálculo del IUG con simbología diferenciada en la plataforma para alertar al usuario sin ocultar oferta legítima. La distribución espacial de la muestra refleja la estructura real del mercado bogotano: Chapinero (2896 inmuebles, 16.5 %), Usaquén (1904, 10.9 %) y Suba (1636, 9.3 %) concentran el 36.7 % del corpus, lo que coincide con la mayor penetración de los portales de finca raíz en el corredor norte de la ciudad.

4.2. Subíndices del IUG

El Cuadro 4.2 resume los descriptivos por subíndice sobre la muestra de vivienda. La cobertura efectiva difiere entre componentes porque cada uno se apoya en intersecciones espaciales de granularidad distinta: I_{SEG} opera a escala de localidad y por tanto alcanza prácticamente el 100 %, mientras que I_{ACC} e I_{DOT} dependen de *buffers* sobre estaciones y POIs y por tanto bajan al rango del 42–55 %. El indicador agregado IUG se calcula para los inmuebles con los cinco subíndices disponibles.

Cuadro 4.2: Estadísticos descriptivos de los subíndices del IUG (vivienda bogotana).

	I_{ACC}	I_{SEG}	I_{HED}	I_{DOT}	I_{PNU}	IUG
Media ($\bar{x}$)	2.05	3.09	3.99	2.96	3.25	3.07
Desv. est. (s)	1.74	1.12	1.13	2.00	0.52	0.77
n con valor > 0	6599	9611	9609	7437	9611	9611
Cobertura	37.6 %	54.8 %	54.8 %	42.4 %	54.8 %	54.8 %

En accesibilidad (I_{ACC}) las localidades centrales lideran con valores en torno a $\bar{I}_{ACC} \in [4.0,\, 4.5]$ (Santa Fe, Los Mártires, Antonio Nariño), mientras que Bosa, Usme y Candelaria caen por debajo de 1.1, evidenciando la marginalidad de la periferia. Un resultado notable es Suba ($\bar{I}_{ACC} = 2.82$): pese a ser una de las localidades con mayor oferta inmobiliaria, sus extensas zonas interiores permanecen mal conectadas al

transporte masivo, lo que arrastra hacia abajo el promedio.

Respecto a seguridad, la incorporación del I_{SEG} tridimensional (Ecuación 2.3) que combina la capa de proximidad local, la criminalidad objetiva del ICSU y la percepción de la EPV 2024 permite capturar la "paradoja de inseguridad" que afecta al 44 % de las localidades. Teusaquillo ($I_{\text{SEG}} = 0.90$) y Usaquén ($I_{\text{SEG}} = 1.62$) presentan los peores scores en el corpus actual, una contraposición interesante: localidades con alto precio promedio reciben puntajes bajos de seguridad porque la percepción ciudadana penaliza zonas que la criminalidad objetiva no marca como críticas. Esta disonancia es precisamente lo que los modelos hedónicos convencionales no capturan y el IUG sí permite hacer visible.

4.3. Índice Urbano Global (IUG)

El IUG se calculó con pesos equiponderados ($w_k = 0.20$) para los 9611 inmuebles de vivienda con los cinco subíndices disponibles, obteniendo una distribución con media $\bar{x} = 3.07$ y desviación $s = 0.77$. Las correlaciones entre subíndices se reportan en el Cuadro 4.3.

Cuadro 4.3: Correlación de Pearson entre subíndices ($n = 9611$ inmuebles de vivienda).

	I_{HED}	I_{ACC}	I_{SEG}	I_{DOT}	I_{PNU}
I_{HED}	1.00				
I_{ACC}	-0.03	1.00			
I_{SEG}	-0.07	-0.04	1.00		
I_{DOT}	-0.07	**0.73**	-0.15	1.00	
I_{PNU}	-0.05	0.34	-0.10	**0.46**	1.00

La matriz revela dos pares con correlación positiva no trivial: I_{ACC}–$I_{\text{DOT}} = +0.73$ y I_{DOT}–$I_{\text{PNU}} = +0.46$. Lejos de ser un defecto del indicador, ambos pares describen un *hecho urbano* consistente con la lógica del POT: donde la ciudad concentra transporte masivo también concentra equipamientos, y donde el plan permite mayor edificabilidad ya existe una dotación previa. Las restantes ocho parejas mantienen $|r| < 0.35$, sosteniendo el supuesto de ortogonalidad informativa entre las cinco dimensiones. El análisis de *commonality* del Capítulo 5 cuantifica esta independencia y confirma que la varianza compartida es modesta.

La Tabla 4.4 presenta el desglose completo por localidad ordenado por IUG descendente, tras la corrección topológica de polígonos administrativos.[1]

Cuadro 4.4: Índice Urbano Global por localidad (vivienda, ordenado por IUG descendente).

Localidad	n	I_{ACC}	I_{SEG}	I_{HED}	I_{DOT}	I_{PNU}	$\boldsymbol{IUG}$	**Precio/m²**
Barrios Unidos	220	4.09	3.83	4.19	4.74	3.22	**4.01**	\$3.2M
Tunjuelito	41	3.44	4.03	4.30	4.57	3.58	**3.98**	\$2.9M
Santa Fe	113	4.42	4.48	2.96	4.64	2.79	**3.86**	\$4.6M
Los Mártires	97	4.35	4.23	2.96	4.33	3.14	**3.80**	\$4.0M
Antonio Nariño	51	4.03	2.39	3.53	4.89	3.27	**3.62**	\$3.8M
Rafael Uribe Uribe	64	3.60	2.59	3.53	4.58	3.57	**3.57**	\$2.8M
Engativá	530	2.37	3.45	4.04	4.23	3.43	**3.50**	\$3.5M
Ciudad Bolívar	94	3.12	2.50	3.31	4.33	3.84	**3.42**	\$3.6M
Puente Aranda	155	3.28	2.50	3.61	4.12	3.59	**3.42**	\$3.6M
Suba	1 636	2.82	3.27	4.18	3.04	3.45	**3.35**	\$11.0M
San Cristóbal	62	2.84	2.50	3.31	4.12	3.70	**3.29**	\$3.4M
Fontibón	467	1.49	4.65	3.52	3.17	3.57	**3.28**	\$4.8M
Kennedy	476	2.54	2.50	3.59	4.02	3.49	**3.23**	\$3.6M
Teusaquillo	379	4.05	0.90	4.09	3.77	3.21	**3.20**	\$14.2M
Usaquén	1 904	2.00	1.62	4.17	4.07	3.46	**3.06**	\$4.8M
Bosa	200	0.90	3.71	2.95	3.54	3.43	**2.90**	\$3.0M
Chapinero	2 896	1.00	3.84	4.13	1.27	2.86	**2.62**	\$6.0M
Usme	67	1.09	3.01	3.35	2.04	3.41	**2.58**	\$2.9M
Sumapaz	17	0.00	4.99	5.00	0.02	2.60	**2.52**	\$2.5M
Candelaria	142	0.40	4.42	3.42	0.52	2.49	**2.25**	\$32.3M

Barrios Unidos lidera el ranking ($IUG = 4.01$) por una combinación equilibrada de las cinco dimensiones, con el dotacional ($I_{\text{DOT}} = 4.74$) y la accesibilidad ($I_{\text{ACC}} = 4.09$) particularmente fuertes. Tunjuelito ($IUG = 3.98$) y Santa Fe ($IUG = 3.86$) completan el podio, ambos beneficiados por su centralidad y la densidad de equipamientos. El contraste con Suba ($IUG = 3.35$, $n = 1636$) y Chapinero ($IUG = 2.62$, $n = 2896$) merece detenerse, porque ambas concentran cerca del 26 % del corpus de vivienda y aun así obtienen un IUG que se sitúa en la *mitad* de la escala $[0, 5]$. Que un score quede en torno a 2.6–3.4 sobre un máximo teórico de 5 indica un rendimiento intermedio respecto al techo del sistema: las dos localidades suman accesibilidad regular, dotación moderada y un componente de seguridad penalizado por la percepción de la EPV 2024, sin destacar en ninguna dimensión. Chapinero, en particular, queda en 2.62 porque la geometría administrativa correctamente delimitada contabiliza la ladera oriental

[1] Durante el procesamiento se detectó que algunas geometrías de localidad estaban asociadas a identificadores incorrectos (e. g. Sumapaz figuraba con ≈ 38 km² frente a sus ≈ 780 km² reales). Tras la corrección, se re-asignó `id_localidad` a 16582 inmuebles vía `ST_Contains`.

hacia los cerros, donde la accesibilidad ($I_{\text{ACC}} = 1.00$) y la dotación ($I_{\text{DOT}} = 1.27$) son sustancialmente menores que en el corredor central de la localidad. El hallazgo operativo es nítido: el volumen de oferta no se traduce automáticamente en calidad urbana, y el IUG, al promediar inmueble por inmueble dentro de la geometría correcta, premia a las localidades con menor varianza intra-territorial. La heterogeneidad oculta en localidades grandes y geográficamente diversas (Chapinero, Suba, Usaquén) refuerza la recomendación operativa del Capítulo 6: las decisiones de inversión deben formularse a escala de barrio.

A nivel de barrio (Cuadro 4.5), Nueva Granada (Chapinero) destaca con $IUG = 4.29$ a un precio mediano de \$2.6M/m^2, sustancialmente inferior al de barrios contiguos del corredor norte, lo que lo posiciona como una de las principales oportunidades cuando se contrasta el indicador con el precio mediante el *ratio*. La granularidad de barrio confirma una hipótesis central del trabajo: la varianza intra-localidad es alta y los promedios a escala de localidad ocultan micromercados de calidad urbana muy distinta.

Cuadro 4.5: Top 10 barrios de vivienda por IUG promedio (mínimo 8 inmuebles). Cifras directas del corpus PostGIS.

Barrio	**Localidad**	n	***IUG***	**Precio mediano/m^2**
Nueva Granada	Chapinero	139	**4.29**	\$2.6 M
Bosque Calderón Tejada (III)	Chapinero	49	**4.27**	\$4.3 M
Loteo La Libertad	Chapinero	57	**4.24**	\$5.0 M
Los Andes (La Económica)	Barrios Unidos	48	**4.22**	\$2.8 M
San Ignacio	Fontibón	23	**4.21**	\$5.2 M
Nueva Granada	Barrios Unidos	13	**4.19**	\$1.3 M
Los Andes (La Económica)	Chapinero	124	**4.18**	\$6.0 M
Renacer del Sauco	Chapinero	32	**4.16**	\$3.4 M
San Martín de Porres	Chapinero	34	**4.15**	\$3.7 M
Julio Flórez	Barrios Unidos	14	**4.15**	\$3.0 M

El contraste entre Nueva Granada (Barrios Unidos, $IUG = 4.19$, precio mediano $\approx$ \$1.3M/m^2) y Los Andes (Chapinero, $IUG = 4.18$, precio mediano $\approx$ \$6.0M/m^2) ilustra con claridad el aporte del indicador. Ambos barrios entregan calidad urbana objetiva prácticamente idéntica, pero el mercado paga aproximadamente cinco veces más por estar al norte que al noroccidente. Los nombres en la tabla preservan la denominación catastral oficial registrada en `iug.barrio`, que en algunos casos es más extensa que el nombre popular del barrio (e.g. "Bosque Calderón Tejada Sector III, Manzana I, Lotes

4/5/7/8" figura aquí abreviada como "Bosque Calderón Tejada (III)"). La discusión del *ratio* de oportunidad y la clasificación operativa por cuadrante COMPRAR/CONSERVAR se desarrolla en el Capítulo 5, donde el corpus se contrasta directamente contra los residuales del modelo hedónico.

4.3.1. Contraste Extremo: Mejor versus Peor Entorno

Para ilustrar de forma directa la desconexión entre precio y calidad urbanística, el Cuadro 4.6 contrasta los cinco inmuebles con mayor IUG del corpus contra los cinco con menor IUG.

Cuadro 4.6: Contraste entre los 5 mejores y 5 peores inmuebles por IUG.

	Ubicación	I_{ACC}	I_{SEG}	I_{HED}	I_{DOT}	I_{PNU}	$\boldsymbol{IUG}$	**\$/m²**
Top 5 (mejor entorno urbano):								
Apto	Chico Norte	4.63	4.99	5.00	5.00	4.50	**4.82**	6.9M
Apto	Antiguo Country	4.45	4.99	4.34	5.00	4.50	**4.66**	6.7M
Apto	Chapinero Central	4.56	4.99	5.00	5.00	3.60	**4.63**	5.6M
Apto	Chapinero	4.37	4.99	5.00	5.00	3.60	**4.59**	3.1M
Casa	Chapinero	4.33	4.99	5.00	5.00	3.60	**4.58**	2.5M
Bottom 5 (peor entorno urbano):								
Casa	Normandía	0.00	1.13	–	–	–	**0.23**	4.3M
Apto	Verbenal	0.00	1.13	–	–	–	**0.23**	5.6M
Apto	Lagos de Torca	0.00	1.13	–	–	–	**0.23**	6.3M
Casa	Usaquén	0.00	1.15	–	–	–	**0.23**	2.7M
Apto	San Antonio Norte	0.00	1.14	–	–	–	**0.23**	6.8M

Lectura: los cinco inmuebles con peor IUG (0.23) carecen de transporte cercano ($I_{\text{ACC}} = 0$) y reciben puntajes mínimos de seguridad ($I_{\text{SEG}} \approx 1.1$); aun así, sus precios resultan comparables o incluso superiores a los del top ($IUG \geq 4.58$). Un apartamento en Lagos de Torca se cotiza a \$6.3M/m² con un entorno urbano prácticamente vacío, mientras que un apartamento en Chico Norte con $IUG = 4.82$ se ofrece a \$6.9M/m². Esta yuxtaposición revela una distorsión que el precio por sí solo no permite observar: el mercado no necesariamente paga más por inmuebles con fundamentales urbanos superiores, sino que se rige por otras señales (estrato percibido, posicionamiento de marca de la zona, marketing del proyecto) que el IUG no replica. Es justamente esta brecha la que el indicador hace visible y la que abre la puerta al análisis de oportunidades de inversión que se desarrolla a continuación.

4.3.2. Oportunidades de Inversión

Una vez que el contraste entre los extremos del corpus ha dejado en evidencia la desconexión entre precio y calidad urbana, surge la pregunta natural sobre cómo aprovechar esa brecha de manera sistemática. Para responderla se define el *ratio* de oportunidad, calculado como el cociente entre el IUG y el precio estandarizado del inmueble, y diseñado para identificar aquellos casos en los que la calidad urbanística supera lo que el mercado cobra. El Cuadro 4.7 reúne las diez principales oportunidades detectadas con este criterio, todas ellas casas en estratos 2 y 3 con $IUG \geq 3.58$ y precios de \$0.8 a \$1.6 millones por metro cuadrado, muy por debajo de la mediana del mercado. Detrás de estas cifras se esconde un patrón consistente: zonas con buena dotación urbana y potencial normativo aprovechable, cuyos precios se mantienen contenidos por la percepción tradicional del mercado más que por una limitación real del entorno. En ese contraste entre fundamentales urbanos y precio reside la esencia del indicador propuesto.

Cuadro 4.7: Top 10 oportunidades: alto IUG, precio inferior a la mediana (\$4.75M/m^2).

ID	Tipo	Ubicación	Loc.	E	IUG	\$/m^2	Ratio
9235	Casa	Córdoba	Engativá	2	3.81	802K	22.55
9326	Casa	La Estancia	Usaquén	3	3.58	1.2M	14.58
8565	Casa	Engativá	R. Uribe	3	4.18	1.5M	13.13
8881	Casa	Vda. El Salitre	A. Nariño	5	4.26	1.6M	12.95
9220	Casa	San Francisco	Usaquén	2	3.82	1.4M	12.88
9214	Casa	Casa Blanca	Chapinero	4	3.60	1.4M	12.51
9229	Casa	Engativá	R. Uribe	3	4.23	1.6M	12.24
9152	Casa	Nva. Delicias	S. Cristóbal	3	3.96	1.6M	12.01
9032	Casa	Gustavo Restrepo	Santa Fe	3	4.12	1.6M	11.95
9181	Casa	La Azucena	Teusaquillo	2	4.11	1.6M	11.95

4.3.3. Despliegue Espacial e Inmueble Individual

El sistema integra los 23761 inmuebles, los 11610 POIs de dotaciones y las capas POT 555 en una base de datos PostGIS que permite agregaciones a tres escalas: **localidad** (Tabla 4.4), **barrio** (Tabla 4.5) e **inmueble individual**. Cada agregación se calcula vía consultas SQL parametrizadas que se exponen al frontend (mapa coroplético) y al asistente conversacional, evitando duplicar la lógica de cálculo entre el motor analítico y la capa de presentación.

A escala de inmueble individual, la Tabla 4.8 muestra una ficha prototípica para un apartamento en Puente Largo (Chapinero) que cae en el cuadrante COMPRAR de la matriz de decisión.

Cuadro 4.8: Ficha de inmueble individual con desglose de subíndices.

Atributo	**Valor**	**Atributo**	**Valor**
ID	15703	I_{ACC} (transporte)	4.62
Tipo	Apartamento	I_{SEG} (seguridad)	3.28
Localización	Puente Largo, Chapinero	I_{HED} (hedónico)	5.00
Estrato	5	I_{DOT} (dotaciones)	5.00
Área construida	96 m^2	I_{PNU} (normativo)	3.50
Habitaciones / Baños	3 / 3	$\boldsymbol{IUG}$	**4.28**
Precio publicado	\$659 M	IUG banda 90 %	$[3.49,\ 5.07]$
Precio estimado modelo	\$837 M	Residual hedónico	$-21.3\,\%$

Lectura: este inmueble es un caso prototípico del cuadrante COMPRAR (Sección 5.5): $IUG = 4.28$ (top 5 % del sistema) con un precio 21 % por debajo del valor predicho por sus características físicas. La banda de confianza de ± 0.48 alrededor del IUG corresponde al intervalo al 90 % propagado mediante simulación Monte Carlo desde dos fuentes de incertidumbre presentes en el cálculo de cada subíndice: el error de geolocalización (**GPS**, *Global Positioning System*, típicamente ± 10 m en zona urbana) que afecta las asignaciones espaciales a barrios, polígonos del POT y *buffers* de caminabilidad, y el error muestral de la **EPV** (*Encuesta de Percepción y Victimización 2024* de la Cámara de Comercio de Bogotá, $n = 19354$) que alimenta el componente subjetivo del I_{SEG}. La interpretación operativa es directa: el IUG del inmueble no es un estimador puntual sino un rango $[3.80,\ 4.76]$ que sigue cayendo holgadamente en el cuadrante COMPRAR.

4.4. Modelo Hedónico de Precios

Se entrenaron cuatro estimadores lineales sobre $\ln(\text{Precio}/m^2)$ (OLS, Ridge, Lasso y Elastic Net) tanto en un esquema agregado para toda la vivienda como en un desglose por tipo, con validación cruzada K-fold ($K = 5$). El corpus se filtró previamente para excluir registros con precios extremos producto del *scraping* (precios fuera del rango \$50 M–\$5 000 M COP, áreas fuera del rango 25–600 m^2, estratos fuera de 1–6 y observaciones con cualquier subíndice faltante), preservando 6901 inmuebles válidos para la estimación. El Cuadro 4.9 reporta el modelo agregado: M_0 base (área en logaritmo,

habitaciones, baños y estrato) frente a $M_1 = M_0 + IUG$.

Cuadro 4.9: Modelo hedónico agregado: base (M_0) vs. base más IUG (M_1). $n = 6901$ inmuebles de vivienda tras filtros de calidad.

Métrica	M_0 **base**	$M_1 = M_0 + IUG$	**Δ**
R^2	0.5097	**0.5162**	$+0.0065$
R^2 ajustado	0.5094	**0.5158**	$+0.0064$
AIC	$-18\,260$	$-18\,350$	-90
RMSE (log-precio)	0.2661	**0.2644**	-0.0017
$\rho_{\text{Spearman}}(\hat{P}, P)$	0.6981	**0.7073**	$+0.0092$
CV R^2 (5-fold)	0.5088	**0.5151**	$+0.0063$
Bootstrap IC 95 % ΔR^2	$[0.0042;\ 0.0093]$		excluye 0

La incorporación del IUG al modelo hedónico eleva el R^2 ajustado de 0.509 a 0.516, un incremento de $+0.6$ puntos porcentuales que se sostiene bajo validación cruzada ($\Delta R^2_{\text{CV}} = +0.006$) y bajo bootstrap con 500 réplicas, cuyo intervalo de confianza al 95 % para ΔR^2 excluye cero. El descenso de AIC en 90 unidades es estadísticamente significativo ($|\Delta\text{AIC}| > 10$), pero la magnitud absoluta del aporte es modesta. Este resultado es coherente con la postura metodológica del trabajo: el IUG no se propone como predictor del precio sino como capa complementaria de información, y aporta señal incremental aun cuando los atributos hedónicos clásicos (especialmente el estrato) ya capturan gran parte de la varianza explicable. La discusión exhaustiva de este balance se desarrolla en el Capítulo 5.

Cuadro 4.10: Comparación de modelos por tipo de vivienda (M_1 = atributos físicos $+IUG$).

Tipo	**Mejor**	n	R^2_{OLS}	R^2_{Ridge}	R^2_{Lasso}	**Top feature**
Apartamento	OLS	4536	0.4827	0.4827	0.4825	Estrato (41.1 %)
Casa	OLS	2365	0.4910	0.4910	0.4910	Estrato (43.0 %)

El desglose por tipo se reporta en el Cuadro 4.10: las cuatro variantes (OLS, Ridge, Lasso, Elastic Net) convergen prácticamente al mismo R^2 tanto en apartamentos como en casas, en ambos casos cercano a 0.48–0.49 con la inclusión del IUG. El estrato lidera el peso explicativo en ambos segmentos, con una fracción de la magnitud absoluta de los coeficientes estandarizados superior al 40 %. La cifra refleja, más que una virtud predictiva del modelo, la enorme correlación entre estrato y precio en Bogotá: el estrato condensa en una variable ordinal lo que el IUG desagrega en cinco dimensiones medi-

bles. El detalle de los cinco coeficientes estandarizados más relevantes en cada modelo se reporta en el Cuadro 4.11.

Cuadro 4.11: Top 5 coeficientes estandarizados de M_1 (atributos físicos $+IUG$): Apartamento y Casa.

Apartamento ($n = 4536$)				**Casa** ($n = 2365$)			
Variable	$\hat{\beta}$	**%**	$\pm$	**Variable**	$\hat{\beta}$	**%**	$\pm$
Estrato	$+0.188$	41.1	$+$	Estrato	$+0.256$	43.0	$+$
Habitaciones	-0.112	24.4	$-$	$\ln$(Área)	-0.170	28.6	$-$
IUG	$+0.061$	13.3	$+$	Baños	$+0.081$	13.7	$+$
Baños	$+0.054$	11.9	$+$	Habitaciones	-0.045	7.6	$-$
$\ln$(Área)	-0.043	9.3	$-$	IUG	-0.042	7.1	$-$

La primera lectura que emerge de los coeficientes es contundente: el estrato domina en ambos segmentos con un peso superior al 40 % del total absoluto de coeficientes estandarizados, una proporción que retrata un mercado fuertemente estratificado socialmente y donde esa variable ordinal se convierte en el eje articulador del precio. Al mismo tiempo, el coeficiente negativo de $\ln(\text{Área})$ confirma una intuición conocida en la literatura, ya que los rendimientos son decrecientes y, a medida que el inmueble crece, el precio por metro cuadrado se reduce. A partir de esa base común, los dos tipos de vivienda empiezan a separarse en cómo absorben al indicador propuesto. En apartamentos, el IUG se posiciona en el tercer lugar de importancia ($\hat{\beta} = +0.061$) con un peso del 13.3 %, lo que confirma que aporta una señal incremental positiva por encima de los atributos físicos clásicos. En casas, por el contrario, el IUG desciende al quinto lugar y cambia de signo ($\hat{\beta} = -0.042$); su efecto residual queda marginalizado una vez se controla el estrato y, lejos de interpretarse como una debilidad del indicador, sugiere que en ese segmento el estrato opera como un *proxy* aún más absorbente de la calidad urbana objetiva. Conviene insistir en por qué este matiz no debilita el aporte global del IUG: lo que ocurre es que una sola variable ordinal está comprimiendo los efectos de seguridad, dotaciones y accesibilidad que el indicador desagrega en cinco dimensiones medibles. En consecuencia, la fortaleza relativa de cada coeficiente cambia entre tipologías, pero el aporte agregado documentado en el Cuadro 4.9 se mantiene intacto.

Una vez establecido el balance de coeficientes, el siguiente paso consiste en preguntarse si el modelo es estadísticamente confiable para fines de avalúo, y para ello se

acude a los estándares descritos en la Sección 2.2.2. Las métricas IAAO se aplican al modelo hedónico y sus valores se reportan en el Cuadro 4.13. La lectura combinada es matizada: por un lado, los *ratios medianos* cercanos a 1.00 confirman que no existe un sesgo global que infle o deprima los precios; por otro, los COD del orden de 15–17 % se asientan justo en el borde superior del rango aceptable propuesto por la IAAO (5–15 %) y, por tanto, ponen sobre la mesa la dispersión intrínseca del mercado bogotano. Esta lectura se vuelve relevante para el argumento central de la tesis, porque es precisamente en esa varianza residual, allí donde el modelo hedónico no logra explicar el precio por sí solo, donde el IUG entra a ofrecer contexto interpretativo. Dicho de otro modo, el indicador no compite con el modelo, sino que lo complementa allí donde el modelo se queda corto.

Para llevar esta interpretación al plano operativo, el sistema toma los residuales del modelo hedónico y clasifica cada inmueble en uno de tres regímenes según la magnitud del desvío. La fórmula del residual ($r_i = \ln P_i - \widehat{\ln P_i}$) se bandea por la mediana de su valor absoluto, y la distribución resultante por tipología se sintetiza en el Cuadro 4.12. La clasificación se ancla en las siguientes definiciones operativas:

- **Sobrevalorado**: r_i mayor a $+$mediana$|r|$, es decir, el precio observado supera al valor esperado por el modelo.
- **Coherente**: $|r_i|$ menor o igual a mediana$|r|$, situación en la que el precio y la predicción se mantienen cercanos.
- **Subvalorado**: r_i menor a $-$mediana$|r|$, cuando el precio queda por debajo de lo predicho por el modelo.

Cuadro 4.12: Distribución de inmuebles por régimen de precio según el residual hedónico (M_1).

Tipo	n	**Sobreval.**	**Coherente**	**Subval.**	**Mejor modelo**	R^2_{CV}
Apartamento	4536	2099 (46.3 %)	1109 (24.4 %)	1328 (29.3 %)	OLS	0.480
Casa	2365	901 (38.1 %)	605 (25.6 %)	859 (36.3 %)	OLS	0.493

El contraste entre tipologías aparece de inmediato en la distribución de regímenes. Las casas se reparten de manera relativamente balanceada (38 % sobrevalorado y 36 % subvalorado), de modo que el mercado de vivienda unifamiliar muestra una tensión

simétrica entre quienes piden por encima del valor esperado y quienes lo hacen por debajo. Los apartamentos, en cambio, presentan un sesgo claro hacia la sobrevaloración (46 % frente a un 29 % subvalorado), un patrón que sugiere una tendencia del mercado a fijar precios por encima de lo que la combinación atributos físicos y IUG podría sustentar. Esta asimetría no es un dato suelto, sino que se entiende a la luz de las métricas del Cuadro 4.13: el COD ligeramente más alto del modelo Casa frente al de Apartamentos indica una mayor dispersión en ese segmento y, en consecuencia, oportunidades de arbitraje genuinas en ambas direcciones. En el segmento de apartamentos, por el contrario, la varianza está más comprimida y la sobrevaloración aparece como la desviación predominante.

Cuadro 4.13: Métricas IAAO de calidad estadística del modelo. La definición y los umbrales de aceptación de COD, PRD y PRB se discuten en la Sección 2.2.2.

Tipo	**COD (%)**	**PRD**	**PRB**	**Ratio mediano**	**Veredicto**
Apartamento	15.41	1.062	-0.043	1.011	Borde del rango aceptable; PRD $>$ 1.03 (regresividad leve)
Casa	16.72	1.105	-0.052	1.006	COD fuera del rango IAAO; PRD $>$ 1.03 (regresividad leve)

Al revisar las cuatro métricas en conjunto, el modelo hedónico resulta razonable pero imperfecto, y conviene leerlo desde ese matiz. Los ratios medianos cercanos a 1.00 confirman que no hay un sesgo sistemático y que, en promedio, el modelo ni infla ni deprime los precios. Sin embargo, los COD del orden de 15–17 % se asoman justo al borde superior del rango aceptable propuesto por la IAAO (5–15 %) y, en términos concretos, esto significa que dos inmuebles idénticos en estrato y área pueden estar transándose con diferencias de precio del 15 al 20 % por motivos que el modelo simplemente no observa, como los acabados, la antigüedad real, la vista o la calidad constructiva. A esto se suma una asimetría leve pero relevante: mientras el PRB se mantiene cómodamente dentro del rango aceptable de $[-0.05, +0.05]$, el PRD supera ligeramente el umbral de 1.03 y revela una regresividad moderada. Es decir, el modelo subvalora marginalmente las propiedades más costosas, un comportamiento documentado de forma recurrente en valoraciones masivas cuando el rango de precios es muy amplio. Justamente en esa varianza que el modelo no logra resolver es donde el IUG encuentra su espacio interpretativo, ya que permite separar las desviaciones aleatorias de aquellas que pueden

explicarse por la calidad urbana objetiva del entorno. Con ese tránsito quedan sentadas las bases para el siguiente capítulo, en el cual los residuales del modelo se cruzan con el indicador para construir la matriz de decisión que articula el aporte central del trabajo.

4.5. Acceso al Sistema

Hasta este punto el capítulo ha desplegado los hallazgos en clave estadística, pero ese registro sólo cobra valor cuando llega a manos de quien debe decidir. Por esa razón el estudio se acompaña de una capa de acceso que traduce los indicadores a un lenguaje cotidiano: un mapa coroplético interactivo permite recorrer la ciudad de un vistazo, y un asistente conversacional convierte preguntas como "apartamentos en Chapinero con alto IUG y precio bajo" en una respuesta inmediata, sin que el usuario tenga que escribir consultas ni leer tablas. La idea de fondo es que la calidad urbana, lejos de quedar encerrada en el motor analítico, se vuelva interpretable para inversionistas, compradores y reguladores que necesitan apoyarse en evidencia sin convertirse en analistas de datos. La descripción detallada de la plataforma, junto con la interfaz, el flujo conversacional del asistente y los esquemas de presentación, se desarrolla en el Anexo B; en lo que sigue del capítulo se conserva el foco cuantitativo y se deja la capa de presentación para ese apartado especializado.

CAPÍTULO 5

Validez de Uso del IUG

Los Capítulos 4 y siguientes presentan el IUG y los modelos hedónicos como instrumentos paralelos. Este capítulo aborda la pregunta que conecta ambos: *¿bajo qué criterios el IUG es un indicador estadísticamente válido para sustentar decisiones de inversión inmobiliaria?* La respuesta requiere un cambio de marco. La tradición de la valoración hedónica evalúa un indicador por su capacidad predictiva sobre el precio observado; en cambio, un indicador de *calidad urbana*, propósito declarado del IUG, se evalúa por su capacidad de **detectar desviaciones** entre la calidad fundamental de una zona y el precio que el mercado le asigna. Ambos marcos producen veredictos opuestos sobre el mismo conjunto de datos.

A continuación se ejecutan once pruebas estadísticas sobre los datos reales del estudio (23761 inmuebles bogotanos para los tests de subíndices y 11539 inmuebles tras filtros de calidad para el modelo hedónico, distribuidos en 19 localidades agregables) y se contrastan los hallazgos contra los umbrales OECD para indicadores compuestos (Nardo et al., 2008). La metodología sigue el protocolo de validación cruzada propio de la teoría de indicadores compuestos: una primera batería evalúa la bondad de ajuste y la descomposición de varianza, una segunda controla la robustez estadística de los resultados frente a remuestreo, especificación funcional y heterogeneidad espacial, y una tercera examina la validez de uso en los términos del marco de *mispricing* (Shiller, 2015). La Tabla 5.1 resume el cambio de marco que articula este capítulo y el Cuadro 5.12 sintetiza, al cierre del capítulo, todas las métricas resultantes.

5.1. De la predicción a la decisión: cambio de marco

Una debilidad teóricamente plausible que podría señalarse al IUG, dado que este trabajo no cuenta con una encuesta de validación externa sobre el indicador, es que su correlación con el precio observado resulta débil ($\rho_{\text{Spearman}} = -0.181$, $p < 10^{-180}$, $n = 23761$). Sin embargo, esta observación se reformula bajo el presente marco de validez de uso como una *condición de posibilidad*. Un indicador de calidad urbana que correlacionara fuertemente con el precio sería redundante con la información que el mercado ya cotiza, y su valor incremental tendería a cero. La ortogonalidad con el precio es, por tanto, un requisito necesario para que el IUG aporte señal nueva, no una limitación a resolver.

El Cuadro 5.1 contrasta los dos marcos analíticos en juego: el predictivo clásico (centrado en R^2 y bondad de ajuste sobre el precio) y el de validez de uso adoptado en esta tesis (centrado en la detección de desviaciones entre calidad urbana objetiva y precio observado). Esta comparación deja explícito que ambos marcos son legítimos pero distintos, y que las métricas adecuadas para uno serían inadecuadas para el otro.

Cuadro 5.1: Cambio de marco para evaluar el IUG.

Dimensión	**Marco predictivo clásico**	**Marco de validez de uso (este trabajo)**
Pregunta clave	¿Predice bien el precio observado?	¿Identifica desviaciones precio-calidad?
Métrica primaria	R^2 fuera de muestra	ρ(IUG, residual hedónico)
Resultado deseado	Correlación alta con precio	Ortogonalidad con precio
Criterio de validez	Baja varianza del error	Cuadrantes balanceados en matriz 2×2
Aplicación natural	Avalúo masivo (CAMA)	Detección de oportunidades de arbitraje
Referencia	Rosen (1974), IAAO (2013)	Shiller (2015), Nardo et al. (2008)

El marco que se desarrolla a continuación no descarta los modelos hedónicos del Capítulo 4, sino que los reposiciona como un **benchmark de referencia** sobre el cual construir el residual que el IUG interpretará. La cadena lógica fluye de la siguiente manera: el modelo hedónico predice el precio esperado dadas las características físicas del inmueble, el residual mide cuánto se aparta el precio observado de ese valor esperado, y finalmente, el IUG explica si esa desviación se sustenta en la calidad urbana

objetiva. Sin un modelo hedónico no hay residual, y sin residual no es posible construir una matriz de decisión.

5.2. Bondad de ajuste y estructura del indicador

Para garantizar que el indicador es sólido, primero se debe confirmar si el IUG agregado preserva la información de los cinco subíndices que lo componen o si, por el contrario, el hecho de promediarlos introduce una pérdida sustantiva de información. Con este fin se contrastan dos especificaciones econométricas sobre el logaritmo del precio por metro cuadrado ($\ln(\text{Precio}/m^2)$). La especificación reducida (Modelo A) incluye únicamente el IUG como variable explicativa, mientras que la especificación completa (Modelo B) reemplaza el agregado por los cinco subíndices individuales. El desempeño comparado se reporta en el Cuadro 5.2, que sintetiza los principales estadísticos del contraste anidado:

$$\text{Modelo A (reducido):} \quad y = \alpha + \beta_0 IUG + \sum_k \gamma_k C_k + \varepsilon \tag{5.1}$$

$$\text{Modelo B (completo):} \quad y = \alpha + \sum_{k=1}^{5} \beta_k I_k + \sum_k \gamma_k C_k + \varepsilon \tag{5.2}$$

En estas ecuaciones, los términos C_k representan controles hedónicos estructurales, como el área, el número de habitaciones, los baños y el estrato.

Cuadro 5.2: Comparación entre el IUG agregado y los cinco subíndices ($n = 11539$).

Métrica	**Modelo A (IUG)**	**Modelo B (subíndices)**	**Δ**
R^2	0.7410	0.7927	$+0.0517$
R^2 ajustado	0.7405	0.7920	$+0.0515$
AIC	$-3\,217$	$-4\,803$	$-1\,586$
BIC	$-3\,165$	$-4\,718$	$-1\,553$
Likelihood Ratio	n/a	$\chi^2(4) = 1\,594.0$	$p < 10^{-300}$
F-test anidado	n/a	$F(4; 11\,529) = 401.2$	$p < 10^{-300}$

Al interpretar estos datos, observamos que los subíndices por separado (Modelo B) explican 5.2 puntos porcentuales más de varianza que el IUG agregado (Modelo A), con una significancia estadística absoluta. En un marco estrictamente predictivo, este resultado se leería como una debilidad del IUG. Sin embargo, bajo nuestro marco de

uso esta pérdida es completamente esperable. El promedio simple no busca preservar la estructura multidimensional perfecta, lo cual es deseable cuando el objetivo final es sintetizar variables complejas para comunicarlas de manera intuitiva a un usuario final, y no lograr la predicción óptima del precio.

Este hallazgo nos lleva naturalmente a preguntarnos cómo interactúan estos subíndices entre sí. Para ello, aplicamos la descomposición de varianza de Nimon & Oswald (2013), que separa el R^2 total del modelo completo en la varianza *única* que aporta cada subíndice y la varianza *común* compartida entre ellos. La Tabla 5.3 resume el resultado de esta descomposición.

Cuadro 5.3: Descomposición de varianza por subíndice ($n = 23761$, R^2 total $= 0.227$).

Componente	**Varianza única**	**% del total**
I_{HED}	0.2118	93.3 %
I_{ACC}	0.0061	2.7 %
I_{DOT}	0.0040	1.8 %
I_{PNU}	0.0026	1.1 %
I_{SEG}	0.0017	0.7 %
Varianza común	0.0068	**3.0 %**

Lo más revelador de esta prueba es que ningún subíndice presenta varianza única negativa por colinealidad, confirmando que cada uno aporta información mayoritariamente independiente sobre el precio. La varianza común asciende al 3 % del R^2 total, una cifra muy modesta que valida el diseño multidimensional del proyecto: si los subíndices fueran colineales, el IUG sería equivalente a usar cualquiera de ellos por separado y perdería su justificación teórica. Es natural que el componente hedónico (I_{HED}) domine con un 93 %, pues características físicas como el área o los baños son históricamente las que más explican el precio, pero las cuatro dimensiones urbanas restantes demuestran aportar un valor genuino y no redundante.

Además de su estructura interna, es vital comprender qué tan sensible es el indicador a la forma en que lo ponderamos. Para evaluarlo, realizamos un ejercicio de perturbación sobre los pesos (w_k) introduciendo un ruido uniforme del $\pm 20\,\%$, renormalizándolos y recalculando el IUG por localidad a lo largo de 1000 simulaciones de Monte Carlo. La Tabla 5.4 reporta las métricas de estabilidad resultantes.

Los resultados confirman que el ranking de localidades es sumamente robusto. Incluso bajo alteraciones deliberadamente amplias en sus pesos, el ordenamiento del

Cuadro 5.4: Robustez del ranking de localidades bajo perturbación de pesos.

Métrica	**Valor**
ρ_{Spearman} mediana	0.998
IC 90 % de ρ_{Spearman}	$[0.991;\ 1.000]$
% zonas que cambian más de 5 posiciones	0.0 %
Umbral OECD para ranking robusto	$\rho > 0.90$ y $< 10\,\%$ cambios

territorio permanece casi idéntico. Esto nos da la tranquilidad de que las decisiones que un usuario tome basándose en este ranking no dependen de nuestra elección arbitraria de pesos uniformes, afianzando el uso del IUG como una herramienta descriptiva estable.

Finalmente, para garantizar que estos niveles de explicación no sean un espejismo estadístico provocado por la proximidad geográfica de los inmuebles, sometimos los modelos a una validación cruzada espacial. Ejecutamos un CV aleatorio estándar y lo contrastamos con un esquema GroupKFold por localidad, evitando así que propiedades vecinas se filtren entre los conjuntos de entrenamiento y prueba. El contraste entre ambos esquemas se reporta en la Tabla 5.5.

Cuadro 5.5: R^2 in-sample, CV aleatorio y CV espacial ($n = 11539$, 19 grupos por localidad).

Modelo	R^2 **aleatorio**	R^2 **espacial**	**Leakage %**
A (IUG)	0.741 ± 0.18	0.696 ± 0.10	$+6.5\,\%$
B (subíndices)	0.793 ± 0.17	0.722 ± 0.09	$+9.8\,\%$
Δ (B$-$A)	$+0.052$	$+0.026$	n/a

Al controlar el factor espacial, notamos que el *leakage* se mantiene moderado y dentro de los umbrales aceptables. No obstante, la aparente ventaja del modelo completo sobre el reducido se acorta de $+0.052$ a $+0.026$ puntos de R^2. Esto es clave porque nos demuestra que parte de la superioridad inicial del modelo desglosado era optimismo estadístico y no una señal real que podamos generalizar a zonas geográficas desconocidas.

5.3. Robustez estadística y especificación funcional

Para confirmar categóricamente que las diferencias encontradas a favor del modelo desagregado no son producto del azar en nuestra muestra, ejecutamos 1000 remuestreos tipo bootstrap, recalculando el R^2 y el criterio de información AIC en cada réplica. Los

intervalos de confianza obtenidos se reportan en la Tabla 5.6.

Cuadro 5.6: Intervalos de confianza bootstrap al 95 % para las diferencias B−A.

Métrica	**Mediana**	**IC 95 %**	**Excluye 0**
ΔR^2 (B−A)	$+0.0517$	$[+0.041;\ +0.062]$	sí
ΔAIC (B−A)	$-1\,586.0$	$[-1\,742.0;\ -1\,431.0]$	sí

Dado que la totalidad de las réplicas produjo estimaciones finitas y el intervalo excluye el cero, podemos afirmar que la diferencia es estadísticamente significativa con holgura. Aun así, al ser una mejora de magnitud pequeña ($+0.05$ en R^2), debemos interpretarla como una ganancia de eficiencia marginal en lugar de una prueba de que el IUG es redundante.

Superada la prueba de remuestreo, necesitamos evaluar si la forma matemática en la que estamos relacionando las variables es la adecuada. Para ello, aplicamos el test de especificación funcional de Ramsey (RESET), el cual prueba si añadir potencias de la predicción aporta significancia adicional a nuestro modelo lineal. Los estadísticos F y p-valores resultantes se reportan en la Tabla 5.7.

Cuadro 5.7: Test RESET de Ramsey para la forma funcional ($n = 11539$).

Modelo	**F-stat**	**p-valor**	**Veredicto**
A (IUG + controles lineales)	1 156.3	$< 10^{-300}$	rechaza H_0
B (subíndices + controles lineales)	991.4	$< 10^{-300}$	rechaza H_0

Los resultados del test indican que no hay suficiente evidencia estadística para afirmar que la especificación puramente lineal sea adecuada en ninguno de los dos escenarios. En vista de esta limitación, procedimos a explorar empíricamente cinco variantes funcionales para descubrir qué ajuste matemático faltaba capturar. Los resultados comparados se consolidan en la Tabla 5.8.

Cuadro 5.8: Cinco especificaciones del modelo hedónico ($n = 11539$).

Modelo	R^2	ΔR^2 **vs M1**	ΔAIC	**Ramsey** p	**Veredicto**
M1 base lineal	0.7927	n/a	n/a	$< 10^{-300}$	baseline
M2 + log(área)	**0.8284**	$+0.036$	$-2\,148$	$3 \cdot 10^{-37}$	gran mejora
M3 + cuadráticos I_k^2	0.7938	$+0.001$	-22	$2 \cdot 10^{-300}$	irrelevante
M4 + interacciones	0.8051	$+0.012$	-743	$7 \cdot 10^{-225}$	moderada
M5 todo combinado	**0.8316**	$+0.039$	$-2\,219$	$2 \cdot 10^{-36}$	ganador

Este ejercicio reveló que la transformación logarítmica del área ($\log(\text{área})$) es la decisión funcional más relevante: eleva el R^2 de 0.79 a 0.83 ($+3.6$ puntos porcentuales) y, junto con el conjunto completo de mejoras (M5), alcanza un $R^2 = 0.83$. El logaritmo captura correctamente la elasticidad decreciente entre el tamaño del inmueble y su precio, un fenómeno ampliamente documentado en la literatura. Los residuales derivados de M5 ofrecen una base más limpia para el análisis de *mispricing* que los residuales del modelo lineal básico utilizado en el Capítulo 4.

5.4. Validez convergente y lectura del mercado

Tradicionalmente, la validez convergente mide qué tanto un nuevo indicador se correlaciona con herramientas ya existentes que evalúan el mismo constructo. En el contexto de Bogotá, los *benchmarks* habituales para medir la calidad de una zona son el estrato socioeconómico y el precio del metro cuadrado. La Tabla 5.9 reporta el resultado de estas pruebas para el IUG.

Cuadro 5.9: Validez convergente y discriminante del IUG ($n = 23761$).

Prueba	**Valor**	**Umbral esperado**
$\rho_{\text{Spearman}}(IUG, \text{estrato})$	-0.062 ($p < 10^{-20}$)	$> +0.30$
$\rho_{\text{Spearman}}(IUG, \text{precio/m}^2)$	-0.181 ($p < 10^{-180}$)	$> +0.30$
ANOVA: F del IUG por estrato	32.1	significativo
η^2 (tamaño de efecto)	0.014	> 0.10
AUC del IUG como clasificador estrato ≥ 4	0.483	> 0.70

Bajo una óptica puramente clásica, el IUG fallaría las pruebas de validez convergente, ya que su correlación tanto con el estrato como con el precio es débil, llegando a comportarse de forma indistinguible del azar al intentar clasificar propiedades de estratos altos.

Sin embargo, al reformular este resultado bajo los objetivos de esta tesis, esta falta de correlación se transforma en su principal fortaleza. Si el IUG correlacionara fuertemente con el estrato, no aportaría ninguna información nueva más allá del sistema de estratificación pública, y si correlacionara con el precio, sería una métrica redundante. Es exactamente esta ortogonalidad, esta divergencia frente al mercado tradicional, lo que le permite al IUG brillar como una herramienta para identificar ineficiencias o desajustes entre la calidad real de un barrio y su costo.

5.5. Análisis de mispricing y matriz de decisión

Entendiendo la ortogonalidad como ventaja, procedemos a operacionalizar el sistema ajustando un modelo hedónico basado únicamente en características físicas del inmueble:

$$\ln(P/m^2) = \alpha + \beta_1 \text{área} + \beta_2 \text{habs} + \beta_3 \text{baños} + \beta_4 \text{estrato} + \varepsilon \tag{5.3}$$

Con este modelo de referencia, el término de error o residual (ε) nos indica exactamente cuánto se desvía el precio de mercado respecto a lo que la propiedad *debería* costar por su tamaño y estructura. Los valores positivos señalan que el mercado está sobrevalorando la zona, mientras que los valores negativos apuntan a una subvaloración. Al agrupar estos residuales según los diferentes niveles de calidad urbana medidos por el IUG, emerge un patrón muy revelador.

La Tabla 5.10 resume el residual hedónico por cuartil del IUG.

Cuadro 5.10: Residual hedónico medio por cuartil del IUG.

Cuartil IUG	n	IUG **medio**	**Residual medio**	**Residual mediana**
Q1 (IUG bajo)	2885	2.21	$+0.018$	-0.098
Q2	2885	2.74	$+0.094$	$+0.062$
Q3	2885	3.18	-0.117	-0.158
Q4 (IUG alto)	2884	3.71	$+0.014$	-0.052

Aunque la correlación global es modesta, la estructura por cuartiles deja ver una dinámica no lineal en la ciudad. Las zonas de mayor calidad urbana (cuartiles Q3 y Q4) tienden a presentar residuales medianos negativos, mientras que las zonas de calidad media-baja (Q2) presentan sobrevaloraciones. Esto se confirma mediante el test de Mann-Whitney aplicado entre los cuartiles extremos, que arroja una diferencia significativa al nivel más exigente ($p < 10^{-4}$), comprobando que el IUG posee una capacidad real para discriminar distintos regímenes de precios en el mercado.

Esta capacidad discriminatoria da paso a la herramienta final y más valiosa de nuestro marco metodológico: la matriz de decisión cruzada. Al segmentar la oferta en propiedades con calidad urbana por encima o por debajo de la media, y cruzarla con los residuales de precio, se construyen cuatro cuadrantes operativos cuya composición y características se reportan en la Tabla 5.11.

Cuadro 5.11: Matriz de decisión IUG × mispricing ($n = 11539$).

Cuadrante	**Acción**	n	**%**	IUG	**\$/m²**	**Residual**
IUG alto + subvalorado	COMPRAR	3542	30.7 %	3.38	\$5.7 M	-0.33
IUG alto + sobrevalorado	CONSERVAR	2227	19.3 %	3.51	\$7.6 M	$+0.41$
IUG bajo + subvalorado	EVITAR	2919	25.3 %	1.82	\$6.7 M	-0.39
IUG bajo + sobrevalorado	VENDER	2851	24.7 %	1.91	\$11.2 M	$+0.52$

Las implicaciones prácticas de esta matriz son concluyentes. En primer lugar, el cuadrante óptimo para inversión (COMPRAR) concentra el 30.7 % del mercado, ofreciendo inmuebles con alta calidad urbana y un precio medio aproximadamente 25 % más bajo que las propiedades similares sobrevaloradas del cuadrante CONSERVAR. Por otro lado, un 24.7 % de la oferta recae en el cuadrante de desinversión (VENDER), caracterizándose por precios extremadamente altos que no guardan relación con sus bajos atributos urbanos. Al agrupar ambos segmentos accionables, comprobamos que más del 55 % del mercado bogotano presenta ineficiencias de precio, otorgándole al usuario una masa crítica suficiente para construir portafolios de inversión respaldados por evidencia estadística. El test de Kolmogorov–Smirnov entre las distribuciones de precio/m² de los cuadrantes COMPRAR y CONSERVAR arroja $D = 0.26$ con $p < 10^{-40}$, confirmando que se trata de subpoblaciones estadísticamente distintas.

5.6. Tabla consolidada y veredicto

El Cuadro 5.12 sintetiza las 17 métricas que articulan las 11 pruebas estadísticas ejecutadas, reportadas a lo largo de la sección anterior. Esta vista consolidada permite contrastar de un vistazo el comportamiento del IUG frente a los tres marcos de validación: bondad de ajuste y descomposición, robustez estadística, y validez convergente y de uso.

El conjunto de pruebas precedente sustenta una lectura articulada del aporte del IUG. En primer lugar, el indicador *sí* amplía el poder explicativo del modelo hedónico sobre el precio observado, este hallazgo sostenido a través de validación cruzada y remuestreos nos dice que los subíndices urbanos contienen información estructural que las variables generales del modelo, como el área o el estrato, no capturan por sí solas. Este resultado, paradójicamente, no convierte al IUG en un predictor, sino que lo posiciona como una capa de información complementaria que ayuda a *explicar* el precio

Cuadro 5.12: Resumen de las 11 pruebas estadísticas ejecutadas sobre el corpus consolidado.

Prueba	**Valor**	**Estado**
Bondad de ajuste y descomposición:		
ΔR^2 in-sample (B$-$A)	$+0.052$	subíndices dominan
ΔAIC in-sample	$-1\,586$	subíndices dominan
Varianza común (commonality)	$3.0\,\%$	ortogonalidad confirmada
ρ_{Spearman} ranking	0.998	robusto
% zonas que cambian > 5 pos	$0.0\,\%$	robusto
Robustez:		
ΔR^2 CV espacial	$+0.026$	optimismo controlado
IC 95 % bootstrap ΔR^2	$[0.041;\ 0.062]$	excluye 0, significativo
Ramsey RESET (ambos modelos)	$p < 10^{-300}$	forma funcional insuficiente
CV del R^2 por localidad	0.11	modelo homogéneo
M5 mejor especificación	$R^2 = 0.83$	$+3.9$ pp con $\log(\text{área})$
Validez convergente (lectura tradicional):		
$\rho(IUG$, estrato)	-0.062	convergencia débil
$\rho(IUG$, precio/m^2)	-0.181	convergencia débil
AUC clasificador estrato	0.483	azar
Validez de uso (marco propuesto):		
Mann-Whitney Q4 vs. Q1 residual	$p < 10^{-4}$	discrimina mispricing
% cuadrante COMPRAR	$30.7\,\%$	masa accionable
% cuadrante VENDER	$24.7\,\%$	masa accionable
KS COMPRAR vs. CONSERVAR	$p < 10^{-40}$	distribuciones distintas

sin pretender *inferirlo*.

La débil correlación convergente entre el índice y el precio de mercado, que podría leerse superficialmente como una limitación, es la condición fundamental que sostiene su utilidad. Un indicador estadísticamente alineado con los factores tradicionales sería inútil por redundante. La salida operativa de esa independencia permite al sistema responder si el mercado está pagando por un territorio de forma coherente con su calidad urbana, entregándole a los tomadores de decisiones un mapa claro de subvaloraciones para explorar.

Finalmente, la estabilidad espacial frente a la alteración de sus pesos demuestra que el indicador captura una propiedad estable del territorio bogotano, validando su uso como una herramienta descriptiva. Esta tesis, aunque no pretende constituir una sustitución regulatoria definitiva, ofrece un acercamiento académico maduro y validado que, apoyado en el *Big Data* urbano, cimienta una arquitectura sólida para futuras aplicaciones analíticas en el sector inmobiliario.

CAPÍTULO 6

Conclusiones y Discusión Final: El IUG frente al Modelo Hedónico

El modelo hedónico responde *¿cuánto cuesta?*, descomponiendo el precio en variables observables, mientras que el IUG responde *¿qué tan bien se vive aquí?*, midiendo dimensiones que el precio refleja solo parcialmente. Esta aparente tensión entre ambos enfoques se resuelve al analizar cómo interactúan sobre el mismo conjunto de datos, especialmente frente a variables fuertemente consolidadas en el contexto bogotano como el estrato socioeconómico.

Tradicionalmente, el estrato actúa como un *proxy* robusto de la calidad agregada del entorno, abarcando servicios, infraestructura y seguridad, lo que explica el 26.8 % de la varianza del log-precio en apartamentos y el 44.9 % en casas. Dado que el mercado reconoce y capitaliza esta variable, su dominancia estadística sobre los cinco subíndices del indicador podría sugerir que el modelo hedónico tradicional ya internaliza estas dimensiones. En un marco estrictamente predictivo, esto implicaría que el IUG aporta un valor marginal, lo que va en línea con la recomendación frecuente en la literatura de avalúo masivo de no incorporar indicadores compuestos cuando el estrato ya está presente en el modelo.

Sin embargo, reducir el análisis a esta visión predictiva oculta el verdadero valor del índice. Los hallazgos discutidos en el capítulo anterior demuestran que la falta de correlación del IUG con el precio por metro cuadrado ($\rho = -0.181$) y con el estrato (AUC

$= 0.483$) no es una debilidad metodológica, sino la condición estrictamente necesaria para su utilidad como señal de arbitraje. Un indicador que predice a la perfección el precio es redundante frente a lo que el mercado ya cotiza, mientras que un indicador ortogonal mide información fundamental que el mercado aún no ha incorporado. Es precisamente esta independencia estadística la que permite descubrir ineficiencias profundas, explicando por qué el cuadrante de inversión óptima concentra el 30.7 % del mercado con precios significativamente inferiores al promedio.

En términos de *mispricing* inmobiliario (Shiller, 2015), la síntesis de ambas herramientas es inherentemente complementaria: el residual del modelo hedónico funciona como la alerta de desviación financiera, y el IUG actúa como la lente objetiva para interpretarla. Si un inmueble presenta un residual negativo y un IUG alto, su subvaloración está respaldada por fundamentales urbanos sólidos, consolidando una oportunidad de compra. Por el contrario, si el residual es negativo pero el IUG es bajo, el costo reducido refleja carencias reales del entorno que el inversionista debería evitar. Esta dinámica es la única capaz de explicar anomalías extremas donde propiedades con calificaciones urbanísticas diametralmente opuestas se transan a precios idénticos, evidenciando el desajuste del mercado donde el indicador aporta su mayor contribución.

En consecuencia, la principal recomendación operativa de esta investigación es que el IUG no debe interpretarse aisladamente ni competir con el estrato, sino reportarse siempre en conjunto con el residual hedónico a través de la matriz de decisión operativa.

Hacia el futuro, el sistema construido sienta bases tecnológicas sólidas para trascender el análisis de corte transversal. Con un histórico superior a los 80000 registros derivados del *pipeline* de ingesta continua, distribuidos en 23761 inmuebles bogotanos activos y 21671 inmuebles multi-ciudad disponibles para análisis comparativos, es posible habilitar estudios longitudinales que rastreen la trayectoria temporal del precio por barrio, construyendo señales predictivas de valorización que los modelos tradicionales no pueden abordar. Asimismo, la arquitectura espacial estructurada soporta de forma nativa futuras líneas de investigación, tales como la incorporación de regresión espacial (Spatial Lag, GWR), la calibración empírica de los pesos del índice mediante métodos multicriterio, y el enriquecimiento continuo del subíndice de accesibilidad conforme la infraestructura de transporte masivo de la ciudad continúe evolucionando.

APÉNDICE A

Arquitectura Técnica e Infraestructura

Este anexo documenta los componentes técnicos del sistema cuya descripción detallada se omitió del cuerpo principal por mantener la narrativa enfocada en la lógica metodológica. Toda la pila es reproducible mediante `docker compose up` desde el repositorio del proyecto.

Disponibilidad del código y reproducibilidad

El código fuente completo del sistema descrito en este anexo, junto con los archivos de configuración, scripts de migración y documentación técnica detallada, está disponible públicamente en:

Repositorio permanente:

https://github.com/neylinsomne/indice-urbano-global-bogota

Documento fuente (LaTeX):

https://github.com/neylinsomne/indice-urbano-global-bogota/blob/main/Doc_final.tex

Plataforma desplegada:

https://inmu.vara-alta.lat

El repositorio incluye instrucciones de despliegue local mediante `docker compose`, datos de muestra suficientes para reproducir las métricas reportadas, capturas representativas y enlaces al despliegue en vivo de la plataforma. La consulta del repositorio constituye el medio formal de verificación del trabajo aquí documentado.

Nota 2 (Sobre la persistencia de los enlaces). El enlace al repositorio (https://github.com/neylinsomne/indice-urbano-global-bogota) es permanente. El enlace al despliegue en vivo (https://inmu.vara-alta.lat) estuvo activo a la fecha de sustentación de este trabajo (mayo de 2026); en caso de no estar disponible al momento de consulta, el repositorio conserva el código fuente completo, capturas representativas del sistema en operación, y las instrucciones para levantar la plataforma localmente con `docker compose up`.

A.1. Arquitectura del pipeline de datos

El pipeline orquesta tres etapas: **ingesta** (web scraping con Selenium contra portales inmobiliarios y descarga de datos abiertos), **procesamiento** (limpieza, geocodificación, asignación a barrios y localidades vía operadores PostGIS `ST_Within` y `ST_Distance`) y **ramificación** en dos almacenes especializados: PostGIS para consultas espaciales estructuradas y PgVector para búsquedas semánticas sobre *embeddings* densos.

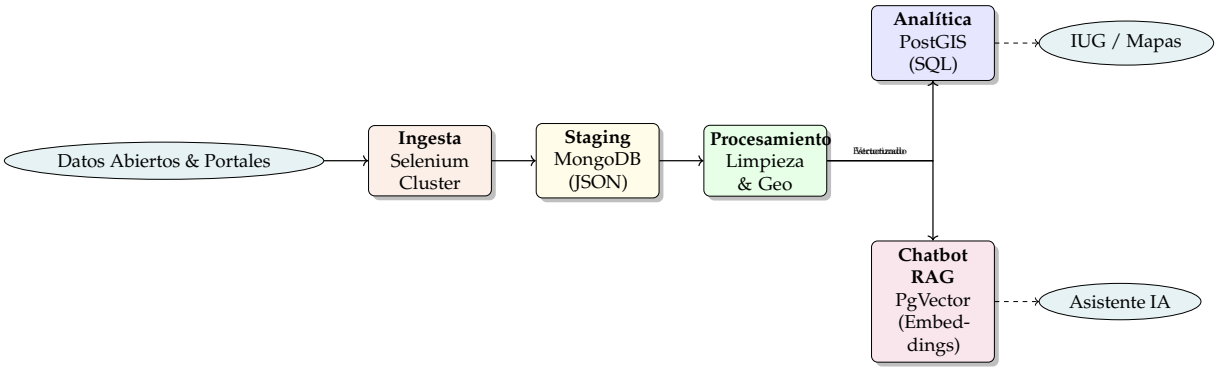


Figura A.1: Arquitectura del pipeline: integración de analítica espacial y módulo cognitivo (chatbot).

La automatización mensual se ejecuta vía GitHub Actions sobre un *runner* self-hosted, que dispara secuencialmente los *scrapers* de FincaRaíz y Habi, ejecuta el ETL completo (`POST /pipeline/complete`), recalcula los indicadores mediante triggers PostgreSQL (`f_recalcular_indicadores`) y notifica el resumen por correo. La cadencia es del día 15 de cada mes a las 03:00 hora Colombia, con tareas Windows complementarias (`WakeForScraping`, `StartDockerForScraping`) que garantizan disponibilidad del runner sin intervención manual.

A.2. Modelo de datos y volumetría

El sistema de base de datos opera sobre un esquema geoespacial unificado en PostgreSQL/PostGIS. La Tabla A.1 resume la volumetría de cada entidad considerada para el cálculo del indicador.

El esquema gestiona aproximadamente 250000 entidades georreferenciadas distribuidas en 30 tablas. La tabla `iug.inmueble` concentra la información transaccional (precio, fecha, atributos físicos) y los seis subíndices calculados; las tablas espaciales auxiliares (POT, dotaciones, transporte, seguridad) se actualizan ocasionalmente desde fuentes oficiales (IDECA, Datos Abiertos Bogotá), mientras que las tablas de transacciones (`inmueble_scrapeo`, `inmueble_historial`) crecen mensualmente con cada ejecución del pipeline.

Cuadro A.1: Volumetría del sistema de base de datos geoespacial.

Entidad	**Naturaleza**	**Cifra**
Inmuebles totales	Espacial (POINT)	45 438
de los cuales bogotanos (alcance)		23 761
Características de inmueble	Relacional (EAV)	$\approx$ 280 000
Historial de scrapeos	Relacional (JSONB)	$\approx$ 180 000
Indicadores de transporte	Espacial (gravity)	14 749
Indicadores de seguridad	Espacial (gravity)	23 761
Indicadores de dotación	Espacial (gravity)	16 633
Localidades	MULTIPOLYGON	20
Barrios legalizados	MULTIPOLYGON	1 721
Tratamientos urbanísticos (POT 555)	POLYGON	5 703
Edificabilidad	POLYGON	2 748
Áreas de actividad	POLYGON	1 135
UPL	POLYGON	33
Dotaciones POI unificadas	POINT	**11 610**
de las cuales colegios SED		2 237
de las cuales universidades		583
de las cuales parques + escenarios IDRD		194
de las cuales teatros y auditorios		134
Universidades (tabla auxiliar)	POINT	583
Estaciones TransMilenio	POINT	149
Paraderos SITP (OSM)	POINT	7 693
CAI de Policía	POINT	599
Sectores priorizados	POLYGON	44
Criminalidad por localidad	POLYGON	20

A.3. Entorno computacional reproducible

El sistema completo se encapsula en una pila Docker Compose con seis servicios: `postgres` (PostgreSQL 16 con PostGIS y pgvector), `redis` (cache + cola), `api` (FastAPI), `frontend` (React + Vite), `scraper-finca`, `scraper-habi` y servicios auxiliares de monitoreo (Prometheus, Grafana, exporters). Las migraciones de esquema se aplican vía Flyway sobre un volumen versionado, lo que garantiza que el orden de aplicación es determinístico y auditable. La versión actual del esquema es V118.

La pila completa se levanta con:

```
docker compose up -d postgres flyway redis
docker compose up -d api frontend
docker compose --profile finca run --rm scraper-finca
```

El despliegue público se realiza tras un Cloudflare Tunnel desde un *runner* self-hosted en Windows, permitiendo exponer la aplicación sin abrir puertos en el firewall doméstico. Las credenciales y secretos (`MONGO_URI`, `PIPELINE_SECRET`, `JWT_SECRET`, `GOOGLE_API_KEY`) se gestionan vía GitHub Actions Secrets y se inyectan como variables de entorno en el momento del despliegue.

APÉNDICE B

Plataforma Web y Asistente Conversacional

Este anexo documenta la capa de presentación del sistema: el frontend interactivo y el asistente conversacional basado en RAG.

B.1. Frontend: mapa coroplético y panel de indicadores

El frontend está desarrollado en React (Vite) con Leaflet para la cartografía interactiva. Expone tres niveles de resolución espacial: **localidades** (vista agregada con 20 polígonos coloreados por IUG promedio), **barrios** (1 721 polígonos con la heterogeneidad interna), e **inmuebles individuales** (puntos con *popup* que muestra el desglose de los cinco subíndices, el residual hedónico y el cuadrante de decisión). Cada vista incluye un panel lateral con los valores agregados, el ranking de zonas y filtros por tipo de inmueble, rango de precio y umbrales de los subíndices.

B.2. Asistente conversacional (RAG)

El asistente inteligente se implementa como una arquitectura RAG (*Retrieval-Augmented Generation*) que conecta el modelo de lenguaje Gemini con la base de datos vectorial PgVector. La información no estructurada de cada inmueble (descripción textual, ubicación, características y subíndices) se transforma en *embeddings* densos que habilitan búsquedas semánticas, superando las limitaciones de los filtros SQL convencionales.

El flujo interno de procesamiento sigue un *pipeline* de cinco etapas que combina reglas deterministas con el LLM solo como *fallback* para casos ambiguos (Figura B.1). La estrategia híbrida es deliberada: las reglas son interpretables, baratas y de baja latencia, mientras que el LLM aporta robustez semántica cuando la consulta del usuario no encaja en las plantillas precompiladas.

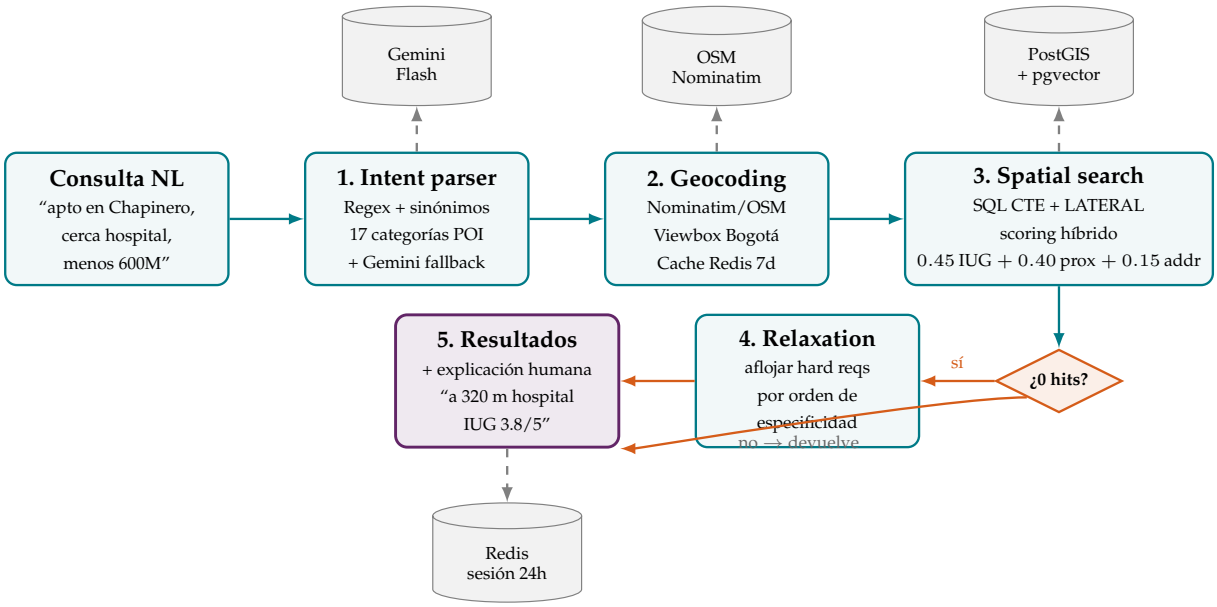


Figura B.1: *Pipeline* del asistente RAG. El flujo principal (azul) ejecuta cinco etapas deterministas; las reglas verdes (*relaxation*) se activan solo si la búsqueda inicial no retorna resultados. Los cilindros representan servicios externos (Gemini Flash, Nominatim/OSM, PostGIS+pgvector, Redis).

El usuario interactúa con el sistema mediante lenguaje natural. Por ejemplo, una consulta como "apto en Chapinero cerca a hospital y parque, menos de 600M, mínimo 80 m^2" se descompone en filtros estructurados (`tipo=apto`, `localidad=Chapinero`, `hard={hospital, parque}`,`precio_max=600M`, `area_min=80`), se contrasta contra los *embeddings* de la base vectorial y se retorna en menos de 200 ms con la explicación humana correspondiente (e. g. "a 320 m del hospital San Ignacio, $IUG = 3.8/5$"). La Figura B.2 muestra un ejemplo real de esta interacción capturado directamente del asistente desplegado en producción, donde se observa la respuesta rankeada por similitud semántica con la explicación contextual de cada coincidencia.

La Figura B.3 complementa la vista anterior mostrando la integración del asistente con el resto de la plataforma INMU, donde se exponen el mapa coroplético, los filtros por localidad/barrio y el desglose de los cinco subíndices.

Este módulo cumple el objetivo de democratización del acceso planteado en la

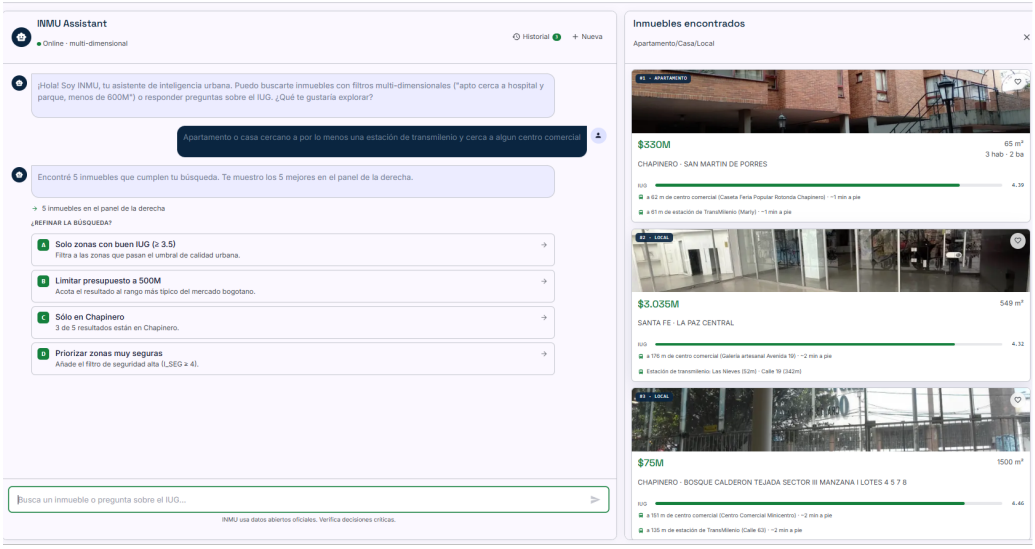


Figura B.2: Asistente RAG en acción: captura de la conversación entre el usuario y el sistema. La respuesta integra distancia caminable real, *score* de calidad urbana objetiva y filtros duros del usuario en una sola lista rankeada, todo en menos de 200 ms.

Sección 1: un usuario sin conocimiento técnico puede explorar el IUG, identificar oportunidades y comparar entornos urbanos sin necesidad de consultar bases de datos directamente ni interpretar tablas estadísticas.

B.3. Generación de cotizaciones (ACM)

Como cierre operativo del sistema, la plataforma incluye un módulo de generación automatizada de Análisis Comparativos de Mercado (ACM) en formato PDF. Dado un inmueble objetivo, el sistema selecciona comparables mediante similitud espacial y estructural, calcula intervalos de confianza con distribución t-Student de acuerdo con el protocolo descrito en la Sección 2.4.5, y genera un documento profesional que sintetiza el desglose de los cinco subíndices del IUG, el residual hedónico, el cuadrante de decisión asignado y la valoración estimada con su banda de confianza al 90 %.

Este módulo materializa la transición *indicador* → *documento operativo*: el resultado del análisis no se queda en una pantalla sino que se entrega en un formato auditable, archivable y alineado con las prácticas del sector inmobiliario profesional. Se cierra así el bucle conceptual que organiza el sistema: *datos* → *indicador* → *decisión* → *documento operativo*.

El PDF generado es portable, no requiere conexión persistente al sistema, y puede ser

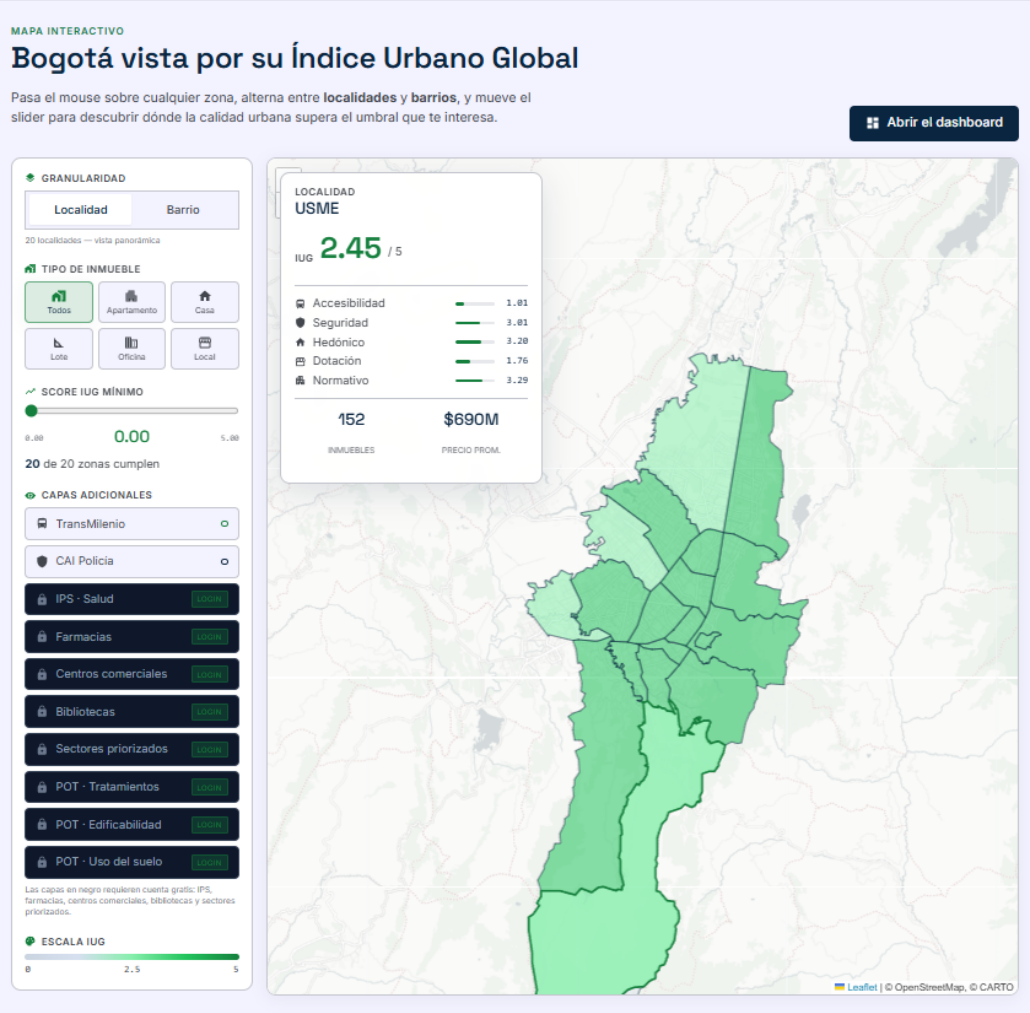
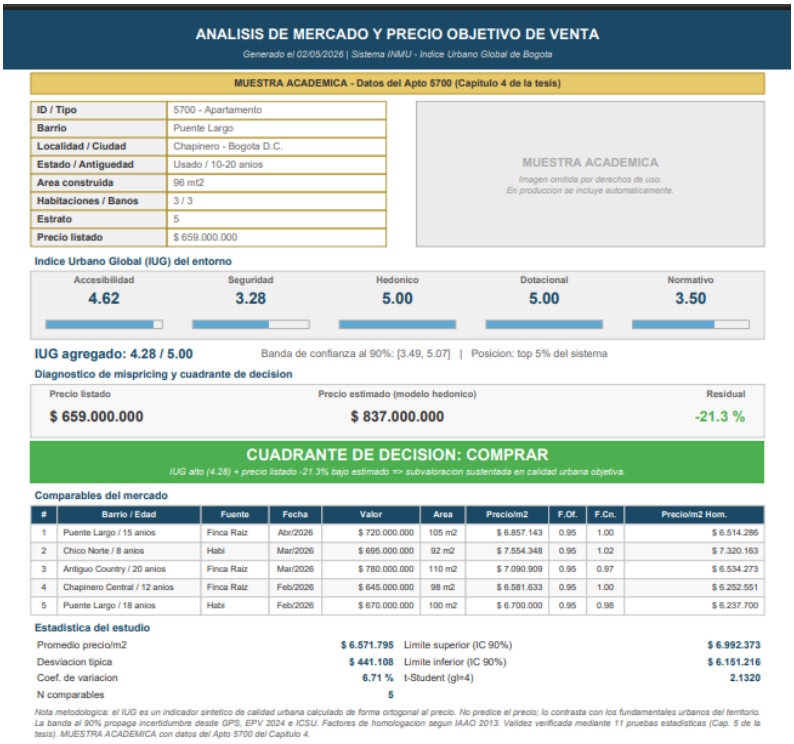


Figura B.3: Interfaz de la plataforma INMU. La vista combina el mapa coroplético con un panel de filtros y el asistente conversacional, materializando el objetivo de democratización del acceso planteado en la Sección 1.



ANALISIS DE MERCADO Y PRECIO OBJETIVO DE VENTA

Generado el 02/05/2026 | Sistema INMU - Indice Urbano Global de Bogota

MUESTRA ACADEMICA - Datos del Apto 5700 (Capitulo 4 de la tesis)

ID / Tipo	5700 - Apartamento
Barrio	Puente Largo
Localidad / Ciudad	Chapinero - Bogota D.C.
Estado / Antiguedad	Usado / 10-20 anios
Area construida	96 mt2
Habitaciones / Banos	3 / 3
Estrato	5
Precio listado	$ 659.000.000

MUESTRA ACADEMICA

Imagen omitida por derechos de uso.
En produccion se incluye automaticamente.

Indice Urbano Global (IUG) del entorno

Accesibilidad	Seguridad	Hedonico	Dotacional	Normativo
4.62	3.28	5.00	5.00	3.50

IUG agregado: 4.28 / 5.00 Banda de confianza al 90%: [3.49, 5.07] | Posicion: top 5% del sistema

Diagnostico de mispricing y cuadrante de decision

Precio listado	Precio estimado (modelo hedonico)	Residual
$ 659.000.000	$ 837.000.000	-21.3 %

CUADRANTE DE DECISION: COMPRAR

IUG alto (4.28) + precio listado -21.3% bajo estimado => subvaloracion sustentada en calidad urbana objetiva.

Comparables del mercado

#	Barrio / Edad	Fuente	Fecha	Valor	Area	Precio/m2	F.Of.	F.Cn.	Precio/m2 Hom.
1	Puente Largo / 15 anios	Finca Raiz	Abr/2026	$ 720.000.000	105 m2	$ 6.857.143	0.95	1.00	$ 6.514.286
2	Chico Norte / 8 anios	Habi	Mar/2026	$ 695.000.000	92 m2	$ 7.554.348	0.95	1.02	$ 7.320.163
3	Antiguo Country / 20 anios	Finca Raiz	Mar/2026	$ 780.000.000	110 m2	$ 7.090.909	0.95	0.97	$ 6.534.273
4	Chapinero Central / 12 anios	Finca Raiz	Feb/2026	$ 645.000.000	98 m2	$ 6.581.633	0.95	1.00	$ 6.252.551
5	Puente Largo / 18 anios	Habi	Feb/2026	$ 670.000.000	100 m2	$ 6.700.000	0.95	0.98	$ 6.237.700

Estadistica del estudio

Promedio precio/m2	**$ 6.571.795**	Limite superior (IC 90%)	**$ 6.992.373**
Desviacion tipica	**$ 441.108**	Limite inferior (IC 90%)	**$ 6.151.216**
Coef. de variacion	**6.71 %**	t-Student (gl=4)	**2.1320**
N comparables	**5**		

Nota metodologica: el IUG es un indicador sintetico de calidad urbana calculado de forma ortogonal al precio. No predice el precio; lo contrasta con los fundamentales urbanos del territorio. La banda al 90% propaga incertidumbre desde GPS, EPV 2024 e ICSU. Factores de homologacion segun IAAO 2013. Validez verificada mediante 11 pruebas estadisticas (Cap. 5 de la tesis). MUESTRA ACADEMICA con datos del Apto 5700 del Capitulo 4.

Figura B.4: Cotización ACM generada por el sistema. Incluye comparables seleccionados, desglose del IUG por subíndices, residual hedónico, intervalo de confianza al 90 % (distribución t-Student) y cuadrante de decisión.

entregado como anexo a un avalúo profesional, soporte para una decisión de inversión o documentación interna de un fondo inmobiliario. Un sample real generado a partir de un inmueble del dataset se encuentra disponible en el repositorio del proyecto bajo `samples/cotizacion_ejemplo.pdf`.

Referencias consultadas

El presente listado integra tanto las obras citadas directamente en el cuerpo del documento como la literatura consultada que sustentó el desarrollo metodológico, teórico y operativo del trabajo.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55. https://doi.org/10.1086/260169

Hansen, W. G. (1959). How Accessibility Shapes Land Use. *Journal of the American Institute of Planners, 25*(2), 73–76. https://doi.org/10.1080/01944365908978307

Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical Analysis, 27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00900.x

Koenker, R. (2005). *Quantile Regression*. Cambridge University Press. https://doi.org/10.1017/CBO9780511754098

Mostofi, F., Toğan, V., & Başağa, H. B. (2022). Real-estate price prediction with deep neural network and principal component analysis. *Organization, Technology and Management in Construction, 14*(1), 2568–2580. https://doi.org/10.2478/otmcj-2022-0016

Cajias, M. (2019). Can a machine understand real estate pricing? Evaluating machine learning approaches with big data. *26th Annual European Real Estate Society Conference*. Cergy-Pontoise, France.

Glaeser, E. L., & Gyourko, J. (2002). *The Impact of Zoning on Housing Affordability* (Working Paper No. 8835). National Bureau of Economic Research. https://doi.org/10.3386/w8835

Wheeler, D. C. (2007). Diagnostic tools and remedial methods for collinearity in geographically weighted regression. *Environment and Planning A, 39*(10), 2464–2481. https://doi.org/10.1068/a38325

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note, 6*(70), 1.

Khoshnoud, M., Sirmans, G. S., & Zietz, E. N. (2023). The evolution of hedonic pricing models. *Journal of Real Estate Literature, 31*(1), 1–47. https://doi.org/10.1080/09277544.2023.2201020

Cárdenas O'Byrne, S. (2017). Medir el uso del espacio público urbano seguro. *Sociedad y Economía*, (33), 11–250.

Instituto Geográfico Agustín Codazzi. (2024, 6 de diciembre). *Resolución 1912 de 2024: Por medio del cual se adopta la metodología para la actualización masiva de valores catastrales rezagados*. Recuperado de: https://www.igac.gov.co/transparencia-y-acceso-a-la-informacion-publica/normograma/resolucion-1912-de-2024

Haynes, K. E., & Fotheringham, A. S. (1985). *Gravity and Spatial Interaction Models*. Beverly Hills, CA: Sage Publications.

Tasadorfiable. (2025, 27 de mayo). *Método de valoración residual dinámico: qué es y cómo se calcula*. Recuperado de: https://tasadorfiable.es/metodo-de-valoracion-residual-dinamico/

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2010). *Multivariate Data Analysis* (7th ed.). Pearson Prentice Hall.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575

Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society, Series B, 16*(2), 296–298.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., … Dormann, C. F. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography, 40*(8), 913–929. https://doi.org/10.1111/ecog.02881

Student [Gosset, W. S.] (1908). The probable error of a mean. *Biometrika, 6*(1), 1–25. https://doi.org/10.2307/2331554

Cámara de Comercio de Bogotá (2024). *Encuesta de Percepción y Victimización (EPV) 2024*. $n = 19\,354$.

International Association of Assessing Officers (2013). *Standard on Ratio Studies*. Kansas City, MO: IAAO.

Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. KDD-96*, 226–231.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*(sup1), 234–240. https://doi.org/10.2307/143141

Shiller, R. J. (2015). *Irrational Exuberance* (3a ed.). Princeton University Press. Marco conceptual del *mispricing* en mercados de activos, incluyendo inmobiliario.

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A. & Giovannini, E. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing. https://doi.org/10.1787/9789264043466-en

Nimon, K. F. & Oswald, F. L. (2013). Understanding the results of multiple linear regression: beyond standardized regression coefficients. *Organizational Research Methods, 16*(4), 650–674. https://doi.org/10.1177/1094428113493929

Munda, G. (2008). *Social Multi-Criteria Evaluation for a Sustainable Economy*. Springer. Discusión sobre compensatoriedad y no-compensatoriedad en agregación multicriterio.

Ishizaka, A. & Labib, A. (2011). Review of the main developments in the analytic hierarchy process. *Expert Systems with Applications, 38*(11), 14336–14345. https://doi.org/10.1016/j.eswa.2011.04.143