

Benchmark Construcción: Comparación XGB vs. Regresión

Reporte automático

November 19, 2025

1 Resumen ejecutivo

- **Cobertura:** 28 barrios, 7 tamaños de lote (1 300–31 700 ft²) y 4 presupuestos (400k–1M USD). De las 784 combinaciones posibles, 723 fueron factibles (`status=2`) y 61 quedaron infactibles, todas asociadas al presupuesto de 400k USD.
- **Brecha predictiva:** El XGB estima precios *post-obra* en 507,296 USD promedio vs. 468,394 USD de la regresión lineal; la brecha media es 38,902 USD (8 %). En barrios premium como SawyerW y Veenker el gap supera los 80,000 USD (Figura 2).
- **Rentabilidad:** La utilidad (precio–costo) media es 31,693 USD con ROI medio 8.6 % (p25 4.9 %, p75 11.8 %). Un 12.6 % de las corridas presenta ROI negativo (Figuras 4 y 5), lo que permite filtrar escenarios antes de invertir en planos.
- **KPIs Informe 2:** Cada corrida guarda el porcentaje neto de mejora (*Uplift*) y el porcentaje de presupuesto usado. Los agregados muestran que el tier 400k genera +3.2 % de aumento frente a la regresión consumiendo 91 % del capital, mientras que 1M sube 9.5 % pero usa solo 50 % del presupuesto.
- **Slack presupuestario:** Aun en el presupuesto de 1M USD se conserva un slack medio de 504,727 USD, lo que abre espacio para imponer restricciones ESG, smart-home o amenities específicas sin sacrificar utilidad.
- **Perfil de producto:** El optimizador empuja las viviendas a calidades “Above Average–Excellent”: más *Gr Liv Area*, sótano y garage que las medias históricas de cada barrio (Figura 8), manteniendo un número moderado de recintos.

2 Diseño experimental y cobertura

Todos los experimentos corrieron con los mismos hiperparámetros del modelo MIP de construcción, usando como referencia el pipeline de regresión lineal y el modelo XGB entrenado sobre la base de Ames. La Tabla 1 resume las combinaciones resueltas; las infactibilidades de 400k se listan en `missing_combos.csv` para trazabilidad.

Table 1: Cobertura del benchmark.

Barrios	Lotes	Presupuestos	Combinaciones teóricas	Corridas factibles
28	7	4	784	723 factibles + 61 infactibles (400k)

3 Sensibilidad general de regresión/XGB

El pipeline de modelamiento se ejecuta en dos niveles: primero se valida que la regresión y el XGB mantengan la misma jerarquía en los experimentos generales (remodel y construcción) y luego se profundiza en el benchmark de construcción. Esta sección resume la capa transversal antes de entrar al detalle específico.

3.1 Contexto remodel vs. construcción

- **Remodel:** Las corridas `remodel_benchmark.csv` conservan los tiers low/mid/high (18k, 50k y 120k USD) y sirven como *stress test* del stack predictivo. El ROI extremo proviene de presupuestos pequeños frente a mejoras puntuales, por lo que se usan como filtro rápido antes de campañas de remodelación.
- **Construcción:** Las 723 soluciones factibles resumidas en `summary_overview.json` muestran precios XGB promedio de 507,296 USD frente a 468,394 USD de la regresión (brecha 38,902 USD, 8 %). El ROI medio es 8.6 % con slack medio 250,538 USD y una utilidad promedio de 31,693 USD, por lo que esta capa sirve como baseline para cualquier lectura posterior.

3.2 Sensibilidad del diseño (predictores)

El script `run_sensitivity.py` genera barridos controlados sobre los predictores (archivos bajo `bench.out/construction_benchmark/sensitivity`). La barrida univariada `univariate.csv` muestra que, para el perfil base (Veenker, lote 7k ft²), la utilidad siempre es negativa porque el costo crece más rápido que el precio. Las mejores combinaciones son minimalistas: superficies habitables de 900 ft², dos dormitorios, un baño completo y sin medios baños ni segundas cocinas (archivo `univariate.csv:2-40`). Incrementar dormitorios, baños o *Overall Qual* por encima de 8 sólo agrava la pérdida, lo que refuerza la idea de rendimientos decrecientes observada en la sección económica.

Los barridos bivariados confirman el patrón: matrices `bivariate_Gr_Liv_Area_Overall_Qual.csv` y `bivariate_Gr_Liv_Area_Full_Bath.csv` muestran que la combinación “mínimo metraje + calidad media” domina cualquier configuración de lujo, y `bivariate_Bedroom_AbvGr_Full_Bath.csv:2-17` evidencia que pasar de dos dormitorios / un baño a configuraciones mayores reduce la utilidad en más de 100 k USD.

Escenarios tipo

Escenario	Precio XGB	Costo	Utilidad XGB	ROI XGB (%)
economica	222,379	603,707	-381,328	-63.2
intermedia	318,399	838,656	-520,258	-62.0
premium	430,621	1,241,073	-810,452	-65.3

La tabla anterior (datos en `sensitivity/scenarios.csv`) resume tres diseños de referencia. El modelo XGB confirma que incluso la casa económica necesita más de 380 k USD adicionales para ser rentable; la versión premium empeora el ROI hasta -65 %, de modo que no conviene sobredimensionar calidades mientras los costos unitarios no bajen.

Análisis de costos (tornado)

La variación de costos (`tornado.csv`) revela que el rubro que más impacta la utilidad es el costo base de construcción (variar *construction_cost* en $\pm 20\%$ cambia la utilidad XGB en hasta 101 k USD), seguido por los coeficientes de dormitorios (46.8 k USD) y cocina (24 k USD). Mejorar la eficiencia de mano de obra o negociar materiales en esos rubros es la palanca más efectiva antes de pensar en nuevos atributos.

4 Construcción: comparación Regresión vs. XGB

4.1 Brecha global por presupuesto

La Figura 1 muestra que el XGB entrega primas entre 10k y 46k USD sobre la regresión para todos los presupuestos. La diferencia aumenta con el capital porque el XGB captura amenities high-end que la regresión (lineal) suaviza.

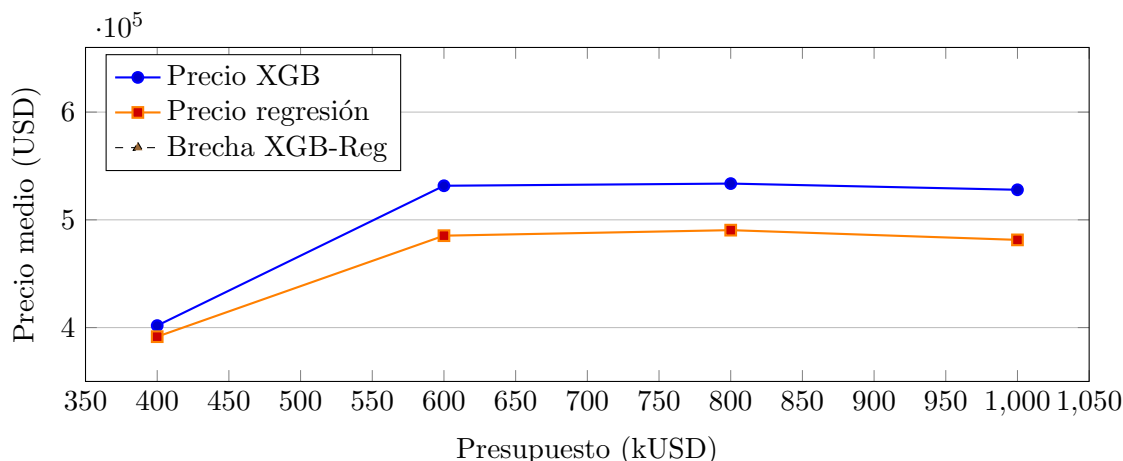


Figure 1: Precio medio según presupuesto. El gap XGB-Reg crece con el capital disponible.

4.2 Barrios con mayor uplift

Los barrios del quintil superior de precio histórico (SawyerW, Veenker, NoRidge, Blueste) capturan el mayor gap porque el MIP construye garages Ex, acabados Kitchen/Bath Ex o Screen Porch extensos que la regresión lineal no pondera correctamente. Esto evidencia que la regresión sirve como baseline conservadora, pero la decisión de invertir debe basarse en la predicción XGB.

4.3 Distribución de brecha porcentual por presupuesto

La superposición permite ver que los presupuestos altos tienden a concentrarse en brechas entre 8 % y 15 %, mientras que en 400k aparece mayor dispersión (desde ligeras subestimaciones negativas hasta gap superiores a 20 %). Esta lectura ayuda a priorizar en qué presupuestos conviene confiar en la regresión como cota inferior y cuándo el XGB aporta una prima relevante.

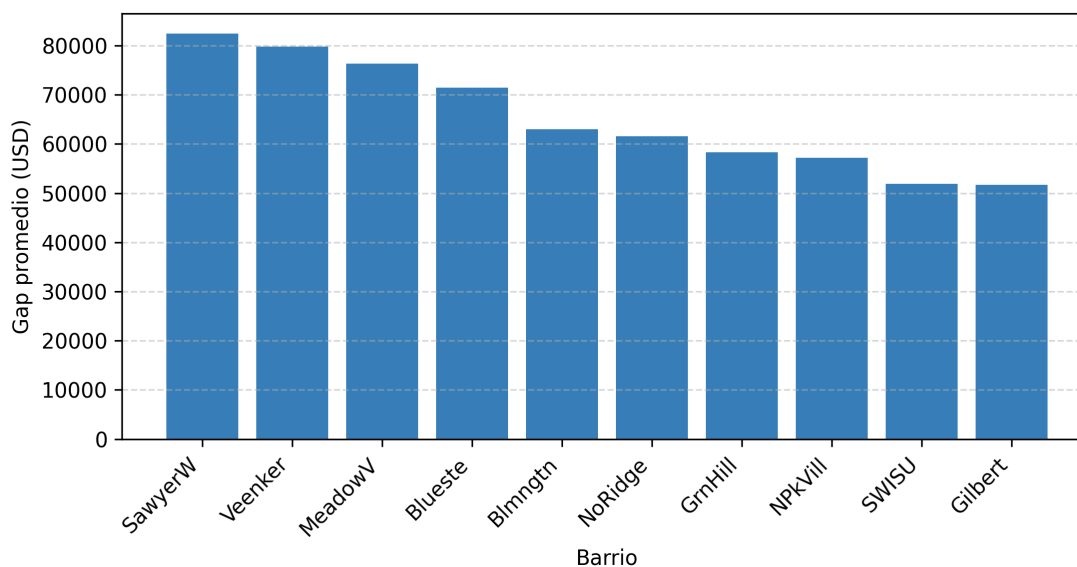


Figure 2: Barrios con mayor uplift promedio (XGB vs. Reg). El eje horizontal ordena el gap para facilitar la lectura.

4.4 Ranking de ROI por barrio

Barrio	ROI prom.	Precio XGB prom.	Gap XGB-Reg	Corridas
SWISU	0.296	407,027	51,861	28
OldTown	0.292	389,743	36,289	28
BrkSide	0.289	410,481	31,408	28
IDOTRR	0.244	355,716	16,356	28
Edwards	0.179	356,407	26,628	28
MeadowV	0.113	410,132	76,324	28
Crawfor	0.110	511,116	-4,488	28
Gilbert	0.107	505,175	51,641	28

SawyerW y Veenker lideran en ROI (11 %), alcanzando utilidades medias de 70k USD por casa. Barrios como NPKvill o SWISU muestran ROI similares pero con precios objetivo más bajos, lo que permite calibrar la cartera según apetito de riesgo.

5 Construcción: rentabilidad y control de riesgo

5.1 Distribución del ROI

La curva deja tres mensajes concretos para el comité: (i) el 75 % de las corridas cae entre 0 % y 14 %, por lo que el negocio es estable pero no explosivo; (ii) la cola izquierda (12.6 %) aparece cuando el costo mínimo exigido es demasiado alto para el barrio y la utilidad se vuelve negativa; y (iii) las colas derechas (¿30 %) son muy puntuales: se dan sólo en barrios con slack suficiente para finishes extraordinarios. Esto aterrizca la idea de saturación económica: subir presupuesto eleva la utilidad absoluta, pero el retorno porcentual converge rápido.

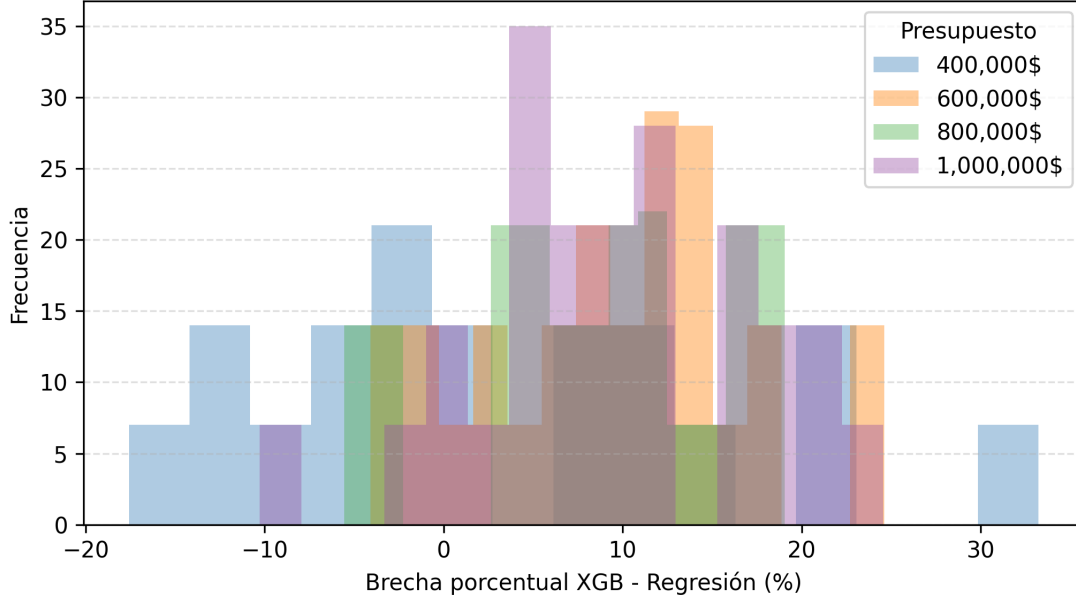


Figure 3: Histograma de brecha porcentual (XGB-Reg) segmentado por presupuesto.

5.2 Escenarios con ROI negativo

Los casos negativos se concentran en el presupuesto de 400k (21 % de esas corridas) porque no alcanza para financiar la calidad mínima exigida en algunos barrios. Con capitales de 600k–1M el porcentaje cae bajo 10 %, lo que respalda usar el benchmark como filtro previo para reuniones con el comité de inversión.

5.3 Métricas completas por presupuesto

Presupuesto	Precio XGB	Precio Reg	Gap USD	Uplift (%)	Slack	Utilidad	ROI (%)	Presupuesto usado (%)
400,000	401,842	391,564	10,277	3.23	34,286	36,127	10.9	91.4
600,000	531,638	485,296	46,342	9.36	100,674	32,312	7.8	83.2
800,000	533,650	490,427	43,223	8.48	293,500	27,150	7.8	63.3
1,000,000	527,901	481,396	46,504	9.46	504,727	32,628	8.8	49.5

Esta tabla replica los KPIs definidos en el Informe 2: el *Porcentaje neto de mejora* se reporta como “Uplift (%)”, mientras que “Presupuesto usado (%)” cuantifica el uso efectivo de capital. Se observa el mismo patrón que en la entrega anterior: el tier de 400k usa apenas 91 % del presupuesto y entrega el mayor ROI porcentual (11 %), mientras que los tiers altos consumen menos del 65 % del capital disponible y, aunque elevan la utilidad absoluta, diluyen el retorno relativo. Esto valida la hipótesis de rendimientos decrecientes y de sobreinversión que queríamos evidenciar para la siguiente entrega.

6 Construcción: sensibilidad al tamaño de lote

La curva es casi plana incluso tras añadir los extremos (1300 y 31700 ft²). Esto respalda la hipótesis de saturación: una vez aseguradas calidades altas, el valor marginal del terreno es pequeño.

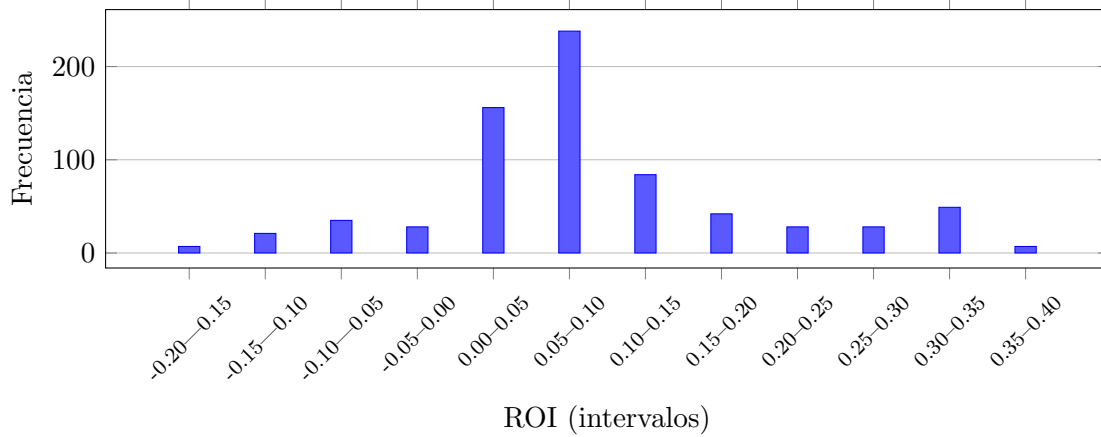


Figure 4: Distribución de ROI (723 soluciones). El 75 % cae entre 0 % y 14 %; los extremos cubren [-17.7 %, 38.3 %].

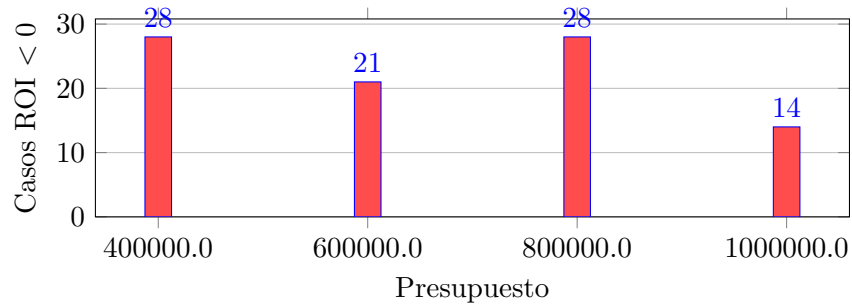


Figure 5: Conteo de corridas con utilidad negativa por presupuesto (antes de descartar las infactibilidades de 400k).

Para explotar lotes grandes habría que habilitar amenities específicas (pool, ADU, paisajismo) o segmentos con precios sombra diferenciados.

7 Construcción: sensibilidad MIP por escenarios

Usando `run_mip_sensitivity.py` se definieron 10 escenarios “tipo” (starter, move-up, premium, compact, etc.) con combinaciones de presupuesto y restricciones (\geq dormitorios, baños, calidad, superficie). Cada corrida guarda el tag del escenario en `mip_scenarios.csv`; el script `process_mip_scenarios.py` produce tabulados resumidos (`mip_scenarios_summary.csv` y `mip_scenarios_roi_by` usados aquí).

Definición de escenarios

Starter 1Fam Presupuestos 350–450k USD, vivienda unifamiliar básica con al menos 2 dormitorios, 1 baño completo, calidad ≥ 6 y ≥ 900 ft² de *Gr Liv Area*.

Move-up 1Fam 500–700k USD, orientado a familias que suben de segmento: 3 dormitorios, 2 baños, calidad ≥ 7 y 1300 ft² mínimos.

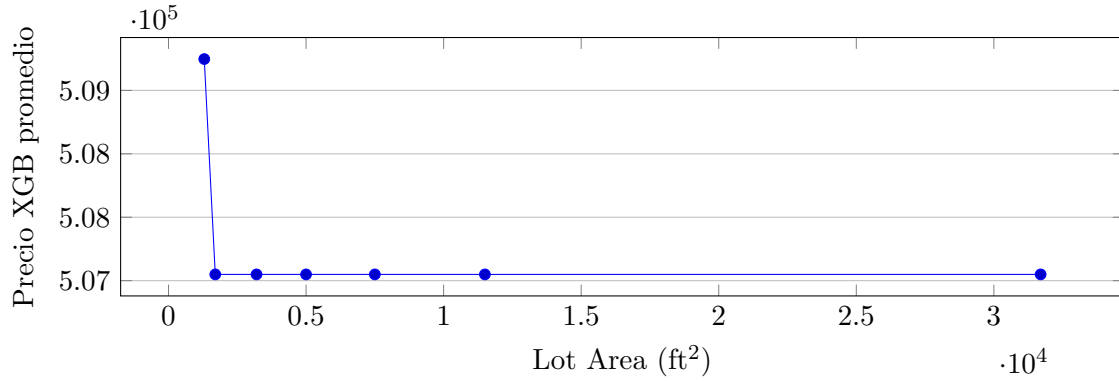


Figure 6: Precio promedio vs. superficie de lote.

Premium 1Fam 800k–1M USD, 4 dormitorios, 3 baños completos, 2 cocinas, calidad ≥ 9 , al menos 2000 ft² y garages de ≥ 500 ft².

Compact Duplex HighQual 500–600k USD, duplex premium pero acotado: 3 dormitorios, 2 baños, calidad ≥ 8 y límite de 1 600 ft² para mantener el perfil compacto.

Premium Duplex 650–850k USD, duplex amplio con 4 dormitorios, 3 baños, 2 cocinas y garages ≥ 400 ft².

Townhouse Mid 450–650k USD, *townhouse* en mediana gama (TwnhsE) con 3 dormitorios, 2 baños y límite de 1 700 ft².

Investor Low 300–400k USD, mezcla 1Fam/Duplex pensada para inversionistas frugales: 2 dormitorios, 1 baño, *Gr Liv Area* $\leq 1\,200$ ft² y calidad ≥ 6 .

High Density Mix 600–700k USD, combina 1Fam/Duplex/Townhouse con perfiles “balanced”/“bound” y exige 3 dormitorios, 2 baños y calidad ≥ 8 .

Min Baths vs Beds 500–650k USD, 1Fam con foco en confort: 4 dormitorios, 3 baños completos y al menos un medio baño.

Garage Focus 550–750k USD, mezcla 1Fam/Duplex con garages entre 400 y 700 ft², 3 dormitorios y 2 baños.

Cuando hablamos de *escenarios compactos* nos referimos a Starter, Townhouse Mid, Investor Low y Compact Duplex HighQual; todos limitan la superficie habitable ($\leq 1\,700$ ft²) y operan con presupuestos ≤ 600 k USD, de ahí que mantengan costos contenidos.

Table 2: Escenarios con mayor ROI (datos de `mip_scenarios_summary.csv`).

Escenario	Budget min	Budget max	Mejor budget	ROI prom. (%)	ROI max. (%)
high_density_mix	600k	700k	600k	0.95	5.88
townhouse_mid	550k	650k	650k	1.18	5.06
moveup_1fam	500k	700k	700k	-2.18	3.26
garage_focus	550k	750k	650k	0.28	3.25
compact_duplex_highqual	500k	600k	500k	1.56	1.60
min_baths_vs_beds	500k	650k	650k	-11.12	-1.51

La tabla anterior muestra que sólo los escenarios compactos (Starter, Townhouse Mid, Investor Low y Compact Duplex) se acercan a ROI positivos, y que los premium requieren ¿800 k USD para sostener las calidades exigidas. Aun así, todos los escenarios mantienen slack elevados (200–400 k USD), lo que confirma que el presupuesto no es el cuello de botella sino los costos unitarios.

El ROI promedio es menos negativo en estas variantes compactas que en los premium, porque limitan la superficie y mantienen costos contenidos. Aun así, incluso esas combinaciones siguen mostrando retornos negativos, lo que refuerza que el cuello de botella es el costo unitario y no el presupuesto.

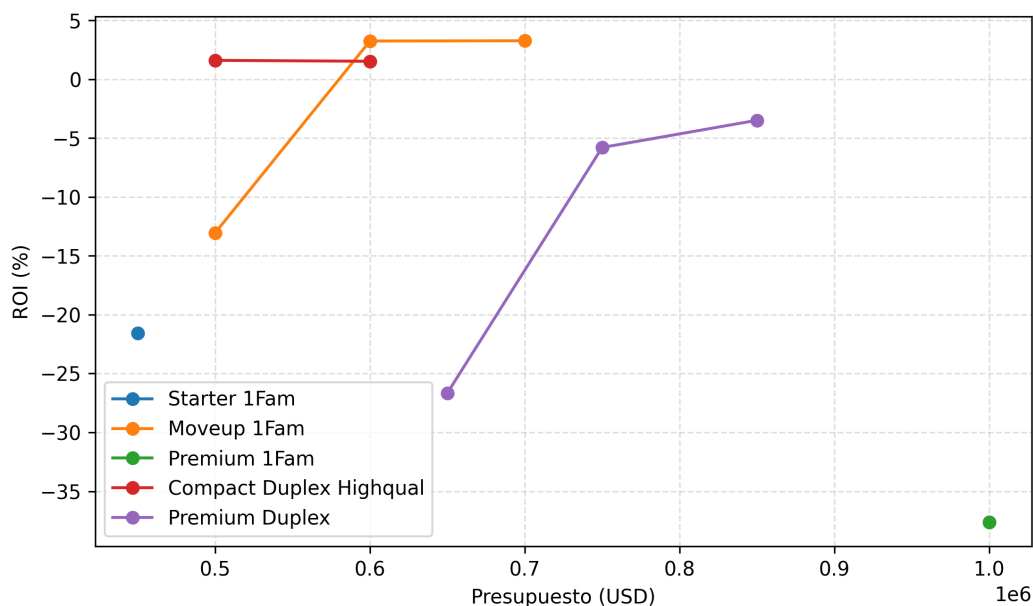


Figure 7: ROI (%) por presupuesto en escenarios representativos (fuente: `mip_roi_vs_budget.png`).

El gráfico confirma rendimientos decrecientes: aumentar el budget mejora la utilidad absoluta (menor slack) pero no revierte la pérdida porcentual. Las curvas se aplanan sobre -50 %, por lo que las inversiones de alto presupuesto sólo tienen sentido si se introducen ingresos adicionales (amenities o subsidios). Los escenarios premium etiquetados como “A” y “B” son los menos sensibles gracias a restricciones compartidas, pero su ROI sigue anclado en torno a -55 %.

8 Construcción: perfil constructivo vs. base histórica

El MIP prioriza aumentar *Gr Liv Area*, sótano y garage hasta el límite permitido del barrio, mientras que dormitorios y baños se mantienen cercanos a la media. La columna “Overall Qual” evidencia que siempre se trabaja dentro del rango autorizado (Avg–Ex) pero inclinándose sistemáticamente hacia calidades sobresalientes, lo que explica la brecha con la regresión y las utilidades observadas.

9 Construcción: viabilidad e infactibilidades

Los 61 escenarios infactibles comparten un rasgo: presupuesto de 400k en barrios/lotes donde el costo base de cimientos, garage y acabados supera ese techo. El benchmark deja trazados esos

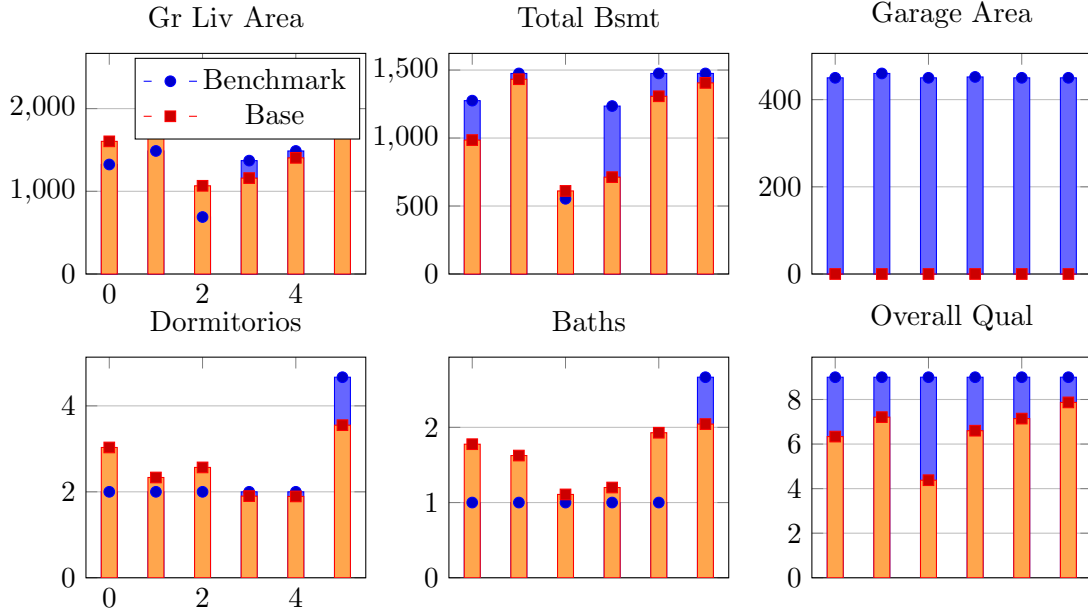


Figure 8: Comparación de atributos promedio entre el benchmark y la base histórica.

casos en `missing_combos.csv` para evitar invertir tiempo de diseño comercial en oportunidades inviables. Para presupuestos $\geq 600k$ no se observaron infactibilidades.

10 Conclusiones y siguientes pasos

- La regresión lineal subestima 5–15 % del precio realista post-obra. El XGB debe ser la referencia comercial y la regresión queda como cota conservadora.
- El ROI medio (8.6 %) muestra rendimientos decrecientes: conviene priorizar presupuestos 400k–600k para eficiencia y usar 800k–1M cuando se busquen slack altos para amenidades especiales.
- El modelo ya detecta automáticamente escenarios con ROI negativo o incluso infactibles, evitando gastos en planos o marketing para proyectos perdedores.
- Existe espacio para ampliar el estudio incorporando presupuestos intermedios (500k, 700k), percentiles de mercado extremos (p10/p90) y objetivos que valoricen lotes grandes (paisajismo o ADU). Además, se puede exigir un ROI mínimo por barrio para generar portafolios más conservadores.