

Residual Constraint Approach (RCA): marco metodológico y protocolo

Traducción al español de la versión actualizada del artículo publicado originalmente en Zenodo el 5 de marzo de 2025. Traducción elaborada con asistencia de IA y revisada por la redacción; la versión en inglés es la autorizada.

William Bert Craytor, SRA

Volumen 1, Número 1 · Verano 2026

Publicado el 1 de julio de 2026

DOI: [10.67330/kgebx46](https://doi.org/10.67330/kgebx46)

© 2026 William Bert Craytor. Published by Pacific Vista Net.

Este artículo de acceso abierto se distribuye bajo los términos de la [licencia Creative Commons Atribución 4.0](#).

Resumen. Una limitación importante de los métodos convencionales de valuación radica en su dependencia de técnicas manuales obsoletas, como el análisis de pares comparables (*matched pairs*), que pasan por alto las contribuciones al valor de las características del inmueble como si no existieran: el *Sales Grid* (cuadro de comparación) tradicional no tiene columna para las contribuciones al valor, solo para ajustes. De hecho, las contribuciones al valor proporcionan la base necesaria para las restricciones matemáticas que impiden la sobrevaluación y la subvaluación por parte de los valuadores tradicionales.

El Residual Constraint Approach (RCA, es.: enfoque por restricciones residuales) hace avanzar el Sales Comparison Approach (SCA, enfoque de comparación de ventas) tradicional mediante la integración de un proceso de valuación multifásico que emplea Multivariate Adaptive Regression Splines (MARS) junto con rigurosas restricciones matemáticas sobre las contribuciones al valor de las características — en particular las características no medidas: variables latentes como la estética, el diseño, el estado y la calidad, cuyos valores suelen estar sujetos al juicio subjetivo y al sesgo del valuador tradicional.

Las restricciones matemáticas del RCA se basan en la aplicación experta de la regresión MARS para estimar las contribuciones al valor de las características medibles del inmueble al precio de venta. Sin embargo, esta estimación normalmente representa solo alrededor del 80 % del precio de venta real, excluyendo características evaluadas subjetivamente como el diseño, la utilidad funcional, el estado y la calidad. En la San Francisco Bay Area, el residuo — que representa estos componentes subjetivos — promedia alrededor del 20 % del valor total del inmueble. La cifra de 80/20 refleja la experiencia del autor en la San Francisco Bay Area bajo condiciones de datos típicas y no debe interpretarse como una constante invariante entre mercados. En mercados donde el R^2 alcanzable es menor, la respuesta del marco metodológico consiste en identificar y medir los determinantes del valor no medidos — la distancia al océano en comunidades costeras, la elevación en vecindarios de ladera y variables similares recopiladas de fuentes externas cuando los datos MLS no las proporcionan — en lugar de aceptar la limitación y reportar bandas de error más amplias. Si bien los residuos de MARS suelen tratarse como errores de estimación en otros contextos, en la valuación inmobiliaria sirven como medida indirecta del valor de las variables latentes. Aunque esto pueda parecer poco práctico al principio, una demostración establece que el residuo puede descomponerse de manera significativa en componentes descriptivos, siempre que sus contribuciones sumen el residuo — sin afectar el resultado final de la valuación.

Esta metodología surge de dos décadas de aplicación empírica por parte del autor, desarrollada mediante implementaciones iterativas de MARS en entornos R y Python.

Palabras clave: valuación inmobiliaria, avalúo residencial, Residual Constraint Approach, MARS, Multivariate Adaptive Regression Splines, Sales Comparison Approach, análisis de residuos, variables latentes, valuación automatizada

Índice

1. Introducción	4
2. Antecedentes	6
3. Valuation Engineer	9
4. Supuestos	10
5. Workflow del RCA	13
5.1. Tipos de análisis MARS	13
5.2. Workflow de procesamiento	13
5.3. Workflow de datos	15
5.3.1. Requisitos de datos para MARS	15
5.3.2. Protocolos: listings de ventas cerradas	16
5.4. Estructura de carpetas del proyecto	18
5.4.1. Protocolos: múltiples ejecuciones de MARS (earth)	19
5.5. El libro de Excel principal	21
5.6. Copias de seguridad	21
6. Notación del libro de Excel	22
6.1. Dimensiones descriptivas	22
6.1.1. La dimensión de proyecto (“a”)	23
6.1.2. Dimensión de versión de ejecución (“r”)	24
6.1.3. Etapas del flujo de trabajo (“w”)	24
6.1.4. Hojas del libro de Excel (“s”)	25
6.1.5. Listings de propiedades (“p”) - filas de MLSDData	25
6.1.6. Variables MLS (“v”) Columnas de MLSDData	26
6.1.7. Ejemplo final de notación	26
6.1.8. Justificación	26
6.1.9. ¿Matrices?	27
6.1.10. Primera simplificación	27
6.1.11. Segunda simplificación	27
7. El Sales Grid del RCA	29

8. Variables	33
8.1. Ontología de origen de las variables	33
8.2. Ontología de aplicación de las variables	34
8.3. Términos de segundo y tercer grado	36
8.3.1. Variables de ajuste	36
8.3.2. Variables de agregación	36
8.3.3. Variables especiales & notación	37
8.3.4. Descomposición de los residuos	37
9. Demostraciones fundamentales del RCA	39
10. Modelo MARS	41
10.1. Funciones “ g ” de Variables del Modelo y de Interacción	41
10.2. Indexación de las Funciones g	41
10.3. Dimensiones de Graficación	42
10.4. La Ecuación Final del Modelo	42
10.5. Número de Gráficas Necesarias para Graficar un Modelo	43
11. Características de los datos	46
11.1. Propósito del residuo	46
11.2. Las calificaciones de estado y calidad como regresores aproximados	47
11.3. Errores de datos	47
11.4. ¿Qué es la exactitud?	49
11.5. El inmueble a valorar como anomalía	50
11.6. La importancia del CVR2 frente a datos inesperados	51
12. Residuos	52
12.1. Ranking & Scoring (clasificación y puntuación)	52
12.2. Curva CQA-Residuo	55
12.3. Características de la curva CQA	55
12.3.1. Análisis de los componentes de la curva CQA	55
12.3.2. Dinámica del mercado e implicaciones sociales	56
12.4. Residuo del inmueble a valorar	56
12.5. Conclusión de valor del RCA	58
12.5.1. Consideraciones generales	58
12.5.2. Revisión, auditoría, quejas	58
13. Conclusión	60

1. Introducción

Este artículo presenta el fundamento lógico y matemático de un enfoque moderno para estimar con exactitud el valor de mercado de inmuebles residenciales y de otros tipos de bienes raíces. El método, desarrollado a lo largo de 20 años de experiencia aplicando el software Multivariate Adaptive Regression Splines (MARS) a viviendas del norte de California — en particular en comunidades del San Francisco Bay Area —, ofrece un avance significativo en la precisión de la valuación. MARS mejora la calidad de la valuación al capturar con destreza las relaciones no lineales entre las características del inmueble y el valor, a la vez que identifica y prioriza de manera eficiente las variables de mayor impacto. Esta capacidad permite a los valuadores descubrir la dinámica del valor en nuevas zonas de mercado con mucha mayor rapidez de la que permiten los métodos tradicionales. MARS constituye la columna vertebral del Residual Constraint Approach (RCA) (es.: enfoque por restricciones residuales) aquí propuesto. El método RCA hace honor a su nombre gracias a su marco riguroso:

- Impone restricciones estrictas sobre las contribuciones al valor de las características del inmueble, tanto medidas como no medidas:
 1. Todas las contribuciones al valor derivadas de MARS, incluido el residuo, deben sumar el precio de venta neto de cada inmueble comparable.
 2. El residuo puede descomponerse en subcomponentes con valores asignados, siempre que su total sea igual al residuo del comparable.
- RCA minimiza el sesgo en la estimación de las características subjetivas — tales como el atractivo estético, el estado y la calidad — al vincular su valor total al residuo, el cual se cuantifica con exactitud cuando un valuation engineer competente construye el modelo MARS.

Aunque está diseñado principalmente para bienes raíces residenciales, donde las características subjetivas influyen de manera significativa en el valor de mercado, RCA tiene potencial para aplicaciones más amplias de valuación de activos. Los avalúos exactos resultan críticos dado el valor estimado de \$45 billones (45 trillion) del mercado residencial de bienes raíces de los EE. UU. Fannie Mae señala que los avalúos típicos se desvían alrededor de un 5 % de los precios de venta reales, aunque esta cifra enmascara el sesgo, ya que los valuadores a menudo enfrentan presiones para alinearse con los precios de venta. Además, dado que el 60 %–70 % de los avalúos respaldan operaciones de refinanciamiento — en las que no existen precios de venta comparables —, la necesidad de precisión, como la que ofrece RCA, se hace evidente. El análisis de la Reserva Federal de 2012 estimó que el 15 %–20 % de los avalúos previos a la crisis de 2008 se desviaban en más de $\pm 10\%$ de los precios de venta, un rango probablemente aún vigente hoy para los avalúos de refinanciamiento, con variaciones según el tipo de inmueble y el mercado. ¿Por qué importa la exactitud? En transacciones con baja relación préstamo-valor (loan-to-value, LTV) — por ejemplo, un comprador con sólido historial crediticio que da \$500,000 de enganche por una vivienda de \$1,000,000 — los prestamistas pueden eximir del avalúo, al suponer un riesgo mínimo de ejecución hipotecaria. Sin embargo, los compradores que omiten el avalúo corren el riesgo de pagar de más y de descubrir defectos solo después de la compra. Esto subraya el valor de métodos de valuación robustos como RCA. Cuantificar la exactitud de los avalúos sigue siendo un desafío. Los investigadores carecen de métricas directas de las desviaciones respecto del “valor de mercado” y recurren, en su lugar, a indicadores indirectos. Los métodos tradicionales de valuación, arraigados en el juicio subjetivo y en la competencia variable de los valuadores, complican el análisis. Entre las estrategias empleadas se encuentran:

1. Tzioumis (2016) estudió a 4,782 valuadores a través de 259,436 avalúos en 11 estados, y encontró que un tercio se ubicaba dentro del 1 % del precio de contrato, la mitad entre el 1 %-3 % por encima, y una minoría en promedio un 5 % por encima — lo que sugiere un sesgo al alza. El estudio identificó dos tipos de valuadores: los “al valor de mercado” (ceranos al precio de contrato) y los “por encima del valor de mercado” (consistentemente más altos).
2. Krivorotov y LaCour-Little (2020) encontró que el 90 % de los avalúos de compra igualaban o superaban los precios de contrato, lo que pone de relieve el “sesgo de anclaje” — la tendencia a alinearse con un valor deseado —, especialmente en contextos de refinanciamiento.

Estos estudios revelan sesgo en las prácticas de valuación, amplificado por las limitaciones de las herramientas tradicionales. Si los valuadores fueran uniformemente competentes e imparciales, sus valuaciones convergerían; la sobrevaluación persistente por parte de ciertos grupos señala problemas sistémicos.

Los estudios sobre sesgo citados anteriormente, y el patrón sistémico que señalan, reflejan un rasgo más profundo de la práctica tradicional de comparación de ventas que RCA está diseñado para abordar. La metodología tradicional opera en dos canales con un estatus epistémico muy distinto. En el canal de las características medibles — GLA (superficie habitable bruta), número de habitaciones, tamaño del lote, antigüedad, ubicación — se espera que el valuador derive los valores de ajuste de manera objetiva, mediante análisis de pares emparejados o regresión lineal local sobre ventas comparables. Los valuadores competentes realizan este trabajo, y los ajustes resultantes son, cuando menos, aproximadamente defendibles frente a los datos. Aun aquí, sin embargo, el Sales Grid tradicional (cuadro de comparación) registra únicamente ajustes; ni siquiera cuenta con una columna para las contribuciones al valor. Tampoco se calculan los ajustes tradicionales como diferencias entre contribuciones al valor — se estiman directamente, a partir de pares emparejados o procedimientos similares, sin que las contribuciones subyacentes lleguen a computarse jamás. En RCA, por el contrario, cada ajuste es la diferencia entre las contribuciones al valor del inmueble a valorar y las del comparable para la característica en cuestión — la estructura cuyas consecuencias precisa la Demostración II. En el canal de las características subjetivas — estado, calidad de construcción, diseño, utilidad funcional, estética — los procedimientos de respaldo disponibles van de endebles a precarios. Los ajustes por características subjetivas pueden respaldarse mediante análisis de pares emparejados de inmuebles ajenos al conjunto inmediato de comparables, mediante apelación a la práctica reconocida entre pares, o mediante referencia a la experiencia documentada del valuador en el mercado. Los tres son formalmente conformes con las normas profesionales. Ninguno alcanza el rigor estadístico que el canal medible logra de manera rutinaria.

El resultado es una asimetría procedimental entre lo que exigen las normas y lo que la metodología puede ofrecer. Las normas exigen respaldo para cada ajuste, sea medible o subjetivo. La metodología ofrece un respaldo robusto en un canal y un espectro de procedimientos más débiles en el otro, que van desde un par emparejado anecdótico de tamaño muestral uno hasta la cita de lo que otros valuadores acostumbra usar. El valuador honesto y competente que realiza un cuidadoso trabajo de pares emparejados para un ajuste por vista produce un respaldo significativamente mejor que el de quien cita la práctica de sus pares sin fundamento; pero un revisor que compare ambos avalúos en el formulario estándar no puede distinguirlos con facilidad, porque el formulario registra el valor del ajuste, no el rigor de su derivación. La metodología tolera la asimetría porque ningún procedimiento más sólido para las características subjetivas ha estado disponible de manera general para los valuadores en ejercicio.

RCA mejora ambos canales, pero de maneras categóricamente distintas. En el canal medible, la regresión MARS captura relaciones no lineales, interacciones entre variables y una estructura de funciones base que ni el análisis de pares emparejados ni la regresión lineal ordinaria pueden

representar. Ajustadas contra el conjunto completo de ventas calificadas de la zona de mercado — típicamente varios cientos — en lugar de contra los tres a nueve comparables del Sales Grid tradicional, las contribuciones al valor derivadas de MARS son a la vez objetivas y sustancialmente más exactas que la metodología a la que reemplazan. En el canal subjetivo, RCA hace algo distinto en especie de cualquiera de los procedimientos tradicionales de respaldo. No intenta en absoluto estimar valores de ajuste para características subjetivas individuales. Las contribuciones de todas las características subjetivas — variables latentes — aparecen colectivamente como el residuo: la diferencia entre el precio de venta neto real de cada comparable y su precio predicho por MARS, calculable a partir de los datos sin intervención del valuador. La descomposición de ese residuo en componentes con nombre — Diseño, Utilidad Funcional, Vista, Estado, Calidad, etcétera — está restringida a sumar el residuo total, y por la Demostración II (Sección 9) cualquier descomposición que satisfaga esa restricción produce la misma conclusión de valor. El juicio que permanece en el flujo de trabajo es la ubicación del inmueble a valorar dentro de la distribución ordenada de los residuos de los comparables, realizada contra una clasificación inspeccionable y documentada explícitamente en el informe. Un avalúo RCA débil sigue siendo posible — un VE inexperto puede producir un modelo MARS sobreajustado, ubicar el inmueble a valorar de manera inverosímil o pasar por alto anomalías en los datos — pero es débil de maneras que un revisor puede identificar al inspeccionar el modelo y la clasificación, y no de maneras que una revisión con el formulario estándar no pueda sacar a la luz.

Más allá del refinanciamiento, los valuadores deben a los propietarios valuaciones exactas, especialmente en compras de alto valor, donde los errores pueden costar decenas o cientos de miles. Sin embargo, muchos se resisten a gastar \$600–\$1,000 en avalúos, dudando de su valor agregado. Los métodos avanzados como RCA, si bien inicialmente intensivos en recursos, prometen mayor precisión. Con refinamiento y experiencia, su eficiencia — y su adopción — deberían aumentar.

2. Antecedentes

La formación académica del autor es en matemáticas, ciencias de la computación y estadística. De profesión, es ingeniero de software con experiencia secundaria en valuación y contabilidad. Ha realizado avalúos, de manera intermitente, desde 2002 y ha aplicado MARS desde aproximadamente 2004 para determinar ajustes de propiedades y tendencias de precios en bienes raíces residenciales.

Valuar con MARS constituye una “experiencia profunda” en valuación. MARS genera modelos que reflejan con mayor detalle que cualquier otro método estadístico tanto las variables que contribuyen al valor de mercado como la manera en que los cambios en dichos valores impactan el precio de venta. Así, al valuar con MARS, el valuador investiga constantemente las relaciones causales entre los valores de las variables y los precios de venta generados por los modelos MARS.

En aras de la transparencia, la metodología RCA (es.: enfoque por restricciones residuales) se aplica mejor a zonas residenciales con las siguientes características:

1. Vecindarios complejos con inmuebles bastante diferentes entre sí

- Viviendas antiguas
- Calidades variadas de renovación
- Muchas arquitecturas y diseños diferentes
- Variación en atributos del inmueble como la GLA (superficie habitable bruta), el tamaño del lote, la elevación y la vista.

2. Condiciones de mercado complejas

- Condiciones de mercado que cambian rápidamente
 - Muchos tipos diferentes de compradores
3. Buenas fuentes de datos: el protocolo RCA se basa en la minería de datos y requiere datos abundantes y de calidad. Si bien puede utilizarse en zonas con datos escasos, podría no justificar el tiempo y el costo requeridos.
 4. Se desean estimaciones exactas del valor de mercado. RCA debería, en la mayoría de los casos, poder proporcionar una exactitud de $\pm 1-2\%$. El informe del valuation engineer debe comentar las condiciones de exactitud en la conclusión de valor final. Si faltan datos o información, o si no se han descubierto patrones consistentes en el mercado mediante la minería de datos, la exactitud se verá afectada y se deberá informar al cliente de la situación. Específicamente, un R^2 (coeficiente de determinación) y un CVR2 (R^2 con validación cruzada, calculado mediante la partición repetida del conjunto de datos de la MLS en particiones aleatorias de entrenamiento y de prueba) elevados, un modelo MARS significativo y un patrón fuerte en las características CQA (estado, calidad y atractivo) del inmueble dentro de los residuos ordenados son los principales medios para evaluar la exactitud y la confiabilidad de la conclusión de valor.

El desarrollo de estos protocolos RCA a lo largo de los últimos más de 20 años se basa en la disponibilidad de MLSListings Inc. (Sunnyvale, CA) y de otras MLS, actualmente accesibles a través de MLSListings, lo cual comprende más del 95 % de la experiencia del autor en análisis de datos. La pericia del autor proviene de un extenso trabajo de valuación en 15 condados primarios del norte de California, con experiencia adicional en 10 condados secundarios.

La metodología RCA es más eficaz en mercados que cumplen dos criterios clave:

1. Amplia disponibilidad de datos de valuación de alta calidad
2. Zonas residenciales complejas y heterogéneas caracterizadas por:
 - Estilos arquitectónicos diversos
 - Grados variables de renovación de los inmuebles
 - Patrones de desarrollo que abarcan varias décadas

Experiencia en condados de California

Las afirmaciones realizadas en este artículo se basan en la experiencia de valuación de inmuebles en los siguientes condados del norte de California:

San Mateo	Santa Clara	Marin	Napa	Monterey	Santa Rosa
Alameda	San Francisco	Contra Costa	Sacramento	El Dorado	San Joaquin
Stanislaus	Placer	Solano	Mendocino	Lake	Colusa
Merced	Nevada	San Benito	Santa Cruz	Sutter	Yuba

Las zonas anteriores abarcan esencialmente el Silicon Valley y sus condados circundantes, extendiéndose hacia zonas más rurales, incluidos el Valle Central y las estribaciones de la Sierra. Cabe mencionar que el autor siempre ha tenido acceso a datos abundantes y *relativamente*

exactos en comparación con lo que muchos valuadores de otras zonas menos desarrolladas de los EE. UU. tienen a su disposición.

Los datos de la MLS utilizados cuentan con certificación RESO y cumplen con el RCF (RESO Common Format). A pesar de la certificación RCF, los datos de importancia usualmente secundaria a veces faltan o son incorrectos para algunos inmuebles, ya sea por errores de captura originales del Tax Assessor (autoridad de valuación fiscal) o por errores posteriores de los agentes de ventas al ingresar datos en la MLS. Con frecuencia estos errores siguen un patrón, como simplemente no realizar una captura si la característica no existe. Por ejemplo, en lugar de ingresar 0 en “carport spaces” cuando no hay cochera abierta, no se ingresa nada. En zonas donde las cocheras abiertas son poco comunes, el analista podría sustituir la ausencia de datos por 0, a fin de obtener un ajuste útil para las cocheras abiertas. Si bien R/earth y otras implementaciones de MARS pueden manejar datos faltantes, corresponde al Valuation Engineer (VE) decidir cómo tratarlos.

Los datos de MLSListings también están disponibles a partir de las MLS de algunos otros estados, desde Florida hasta Hawái. Esto puede ser útil para probar qué tan bien funcionan los protocolos y el código en diferentes estados.

3. Valuation Engineer

Definición 1: Valuation Engineer

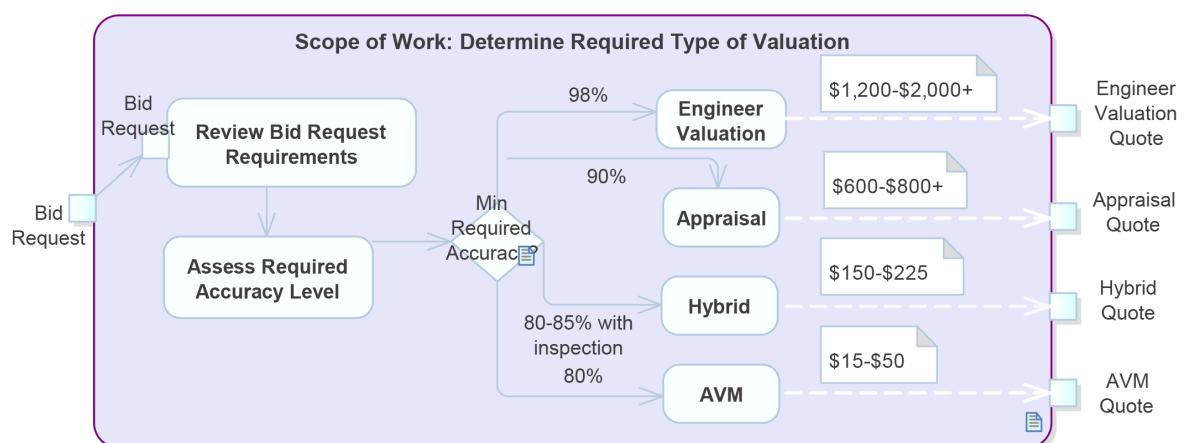
Para evitar ambigüedades, el término “Valuation Engineer” (o “VE”, en forma abreviada) se utiliza para denotar a un individuo que posee la pericia de un valuador con licencia, junto con conocimientos, habilidades y experiencia avanzados en la utilización de minería de datos, estadística, programación, protocolos detallados e inteligencia artificial para determinar el valor de mercado de inmuebles complejos. También es importante reconocer que la valuación es una tarea que requiere alguna forma de licencia debido al potencial comprobado de fraude.

Los prestamistas suelen ofrecer varias opciones de valuación:

1. Modelos de valuación automatizada (AVM, por sus siglas en inglés) - la opción de menor costo
2. Avalúos híbridos - que utilizan inspectores sin licencia bajo la supervisión de un valuador con licencia
3. Avalúos estándar - realizados por valuadores con licencia

Ausentes de la consideración habitual para el trabajo de valuación se encuentran los avalúos de nivel de ingeniería. Sencillamente no hay muchos valuadores con el conjunto de habilidades necesario para realizarlos. Debería existir una necesidad de ellos. Pueden proporcionar una exactitud significativamente mayor y casi garantizar una objetividad bastante buena mediante el empleo de la metodología RCA (MARS y restricciones).

Para inmuebles conformes en fraccionamientos más nuevos, los avalúos estándar generalmente son suficientes. Sin embargo, las valuaciones de nivel de ingeniería se recomiendan para viviendas antiguas en zonas con un historial significativo de renovaciones que abarca más de 30 años. La elección entre estos métodos debe seguir un proceso de decisión estructurado basado en las características del inmueble y las condiciones del mercado.



4. Supuestos

Supuesto 1: Área geográfica & tipo de inmueble

El alcance de este artículo se centra principalmente en viviendas unifamiliares y condominios en las áreas metropolitanas y urbanas de California. Si bien el Sales Comparison Approach (enfoque de comparación de ventas) se aplica a otros tipos de inmuebles y regiones geográficas, los enfoques de ingresos y de costos suelen desempeñar un papel más dominante al valorar terrenos agrícolas, propiedades multifamiliares y bienes raíces comerciales. En tales casos, el enfoque RCA puede requerir modificaciones o tener una aplicabilidad limitada.

Supuesto 2: Almacenamiento de datos

CSV y Excel son ampliamente utilizados por los sistemas MLS y se asumen como la vía principal para alimentar datos a los programas RCA. Aunque los feeds de brokers pueden utilizarse para alimentar datos de manera continua a bases de datos SQL. El respaldo y el almacenamiento intermedio a menudo emplean bases de datos SQL, y dos que desempeñan papeles importantes son las gratuitas y de código abierto SQLite y PostgreSQL. Los archivos de texto o RTF también se utilizan para fragmentos de informes, y los archivos PNG, JPG y TIFF se utilizan para gráficos. Nótese que el SQL Server de Microsoft también es utilizado por muchos, pero la edición gratuita “SQL Server Express” es muy limitada (p. ej., solo se puede usar 1 núcleo de CPU) y las demás versiones tienen cuotas de licencia considerables para uso comercial. Las bases de datos de bienes raíces en California pueden ser muy grandes y los joins pueden consumir mucho tiempo, por lo que una computadora con 8 o más núcleos y 128G de RAM puede ser útil para joins grandes. Las bases de datos SQL también son importantes cuando se necesita encontrar todos los inmuebles dentro de un área delimitada arbitrariamente. Así, si por ejemplo se utilizó el análisis de conglomerados (Cluster Analysis) para definir vecindarios, uno se inclinaría a usar alguna base de datos SQL. Sin embargo, las distintas bases de datos tienen distintas funciones integradas para manejar información geográfica y, por lo tanto, no son compatibles entre sí. Para esta serie de artículos sobre RCA, se asumirá PostgreSQL para las tareas SQL complejas y SQLite para las tareas simples.

Supuesto 3: Se utiliza “R/earth” para el procesamiento MARS.

Hasta 2017, el software MARS de Salford Systems era una solución asequible a \$270 por año. Sin embargo, tras la adquisición de Salford Systems por parte de Minitab, MARS fue incorporado a su paquete SPM a \$16,000 anuales — un precio prohibitivo para la mayoría de las pequeñas empresas. Afortunadamente, el paquete “earth” del lenguaje de programación R (Milborrow [2024a](#)), mantenido por Stephen Milborrow, ha evolucionado hasta igualar o superar las capacidades del SPM MARS de Minitab. El paquete earth ofrece ventajas adicionales: está disponible de forma gratuita, se integra sin problemas con el completo ecosistema estadístico de R y puede accederse desde Python mediante Pandas, lo que proporciona una excelente flexibilidad para diferentes flujos de trabajo. Las discusiones relacionadas con MARS asumen el uso específico de R/earth.

Supuesto 4: Notación

Los datos crudos provenientes de fuentes externas, como archivos CSV y hojas de cálculo de Excel, se representan mediante caracteres latinos estándar. Esto incluye los datos de inmuebles del Multiple Listing Service (MLS), los ajustes de configuración de MARS (Multivariate Adaptive Regression Splines) mediante el paquete R/earth y los datos específicos del proyecto. Esta notación se extiende a sus correspondientes representaciones como data frames en R o Python.

Al tratar las *características del inmueble* y sus diversos tipos y subconjuntos, la notación cambia a letras griegas para una mayor precisión y claridad analítica. Si bien esta notación matemática puede parecer más rigurosa, cumple un propósito crucial: distinguir las características abstractas del inmueble de sus representaciones en datos crudos. Por convención, los caracteres griegos denotan consistentemente características del inmueble, mientras que los caracteres latinos indican estructuras de datos de hoja de cálculo. Por ejemplo, el símbolo M representa la matriz de datos de la MLS, organizada con los inmuebles como filas y sus características correspondientes como columnas.

Supuesto 5: Lenguajes de programación

En este campo, R y Python con Pandas son los lenguajes dominantes, utilizándose C o C++ cuando la velocidad es importante. R/earth está escrito, a su vez, en R y C.

Supuesto 6: Indexación

La indexación puede variar entre los lenguajes de programación y las matemáticas, lo cual puede resultar confuso:

- Las matemáticas y R comienzan la indexación en 1.
- C, C#, C++ y Python comienzan en 0.

En este artículo, para mantener la consistencia y la simplicidad, toda indexación comienza en 0 (como en Python o C). Funciona de la siguiente manera:

- Para listas, matrices o data frames:
 - El índice 0 se reserva para el “inmueble a valorar” (el inmueble principal objeto de estudio).
 - Ejemplo: en una tabla donde las filas son inmuebles, la fila 0 es el inmueble a valorar y la columna 0 contiene los identificadores únicos de todos los inmuebles.
 - Un consejo práctico: ordene los inmuebles con el inmueble a valorar en primer lugar (fila 0) y luego liste los demás de la fecha de venta más reciente a la más antigua, numerándolos secuencialmente (0, 1, 2, etc.).
- Cuando no hay inmueble a valorar:
 - Si una estructura de datos no incluye inmuebles (p. ej., solo variables o funciones), el índice 0 se omite y la indexación comienza en 1.
 - Nota especial para usuarios de R: dado que R utiliza indexación basada en 1, al programar en R las matemáticas de este artículo, sume 1 a todos los índices para ajustar de la indexación basada en 0 a la basada en 1.

5. Workflow del RCA

Comprender el flujo de trabajo del RCA es esencial para un análisis eficaz. El proceso sigue una secuencia estructurada de tareas analíticas y puntos de decisión

5.1. Tipos de análisis MARS

El flujo de trabajo del análisis MARS depende de si se necesita únicamente un análisis de mercado o si se requiere la valuación o el avalúo de un inmueble determinado. Obsérvese que el avalúo de un inmueble específico normalmente requiere un análisis de su zona de mercado.

- **Análisis de mercado:** se centra en áreas amplias tales como condados, ciudades o regiones definidas. Este enfoque no involucra un inmueble a valorar específico y, por lo tanto, no requiere una conclusión de valor. El flujo de trabajo se concentra en el desarrollo del modelo, su validación y la generación de analítica de soporte. El proceso concluye una vez que el modelo óptimo ha sido identificado y documentado.
- **El análisis de un inmueble individual** comprende todos los elementos del análisis de mercado más la determinación del valor de un inmueble específico. Esto difiere de manera fundamental de los métodos tradicionales de valuación. En los enfoques tradicionales, el valor se deriva típicamente ajustando de 3-9 ventas comparables y calculando un promedio ponderado. En cambio, el RCA determina el valor principalmente mediante dos componentes:
 1. El valor predicho por el modelo MARS
 2. El residuo estimado específico del inmueble (con la puntuación CQA asociada)

El Sales Grid en el RCA sirve principalmente para explicar las diferencias de valor entre el inmueble a valorar y los comparables, ilustrando las contribuciones de las variables. Una ventaja clave del RCA es su estabilidad: si bien los clientes pueden debatir la selección de comparables, la conclusión de valor permanece consistente independientemente de cuáles comparables se elijan para la presentación, ya que todas las ventas (sean 15 o 600) se ajustan mediante el mismo modelo. Aunque las ventas más antiguas se ajustan automáticamente a las condiciones actuales del mercado, las ventas recientes son, en general, preferibles. La eficacia del modelo para capturar los factores de precio disminuye con la distancia temporal, pues las características de mercado no medidas cobran mayor importancia en períodos más largos.

5.2. Workflow de procesamiento

El flujo de trabajo general de procesamiento puede describirse mediante estos pasos:

1. Alcance del trabajo.
2. Configurar las carpetas y archivos iniciales del proyecto.
3. Descargar los datos del MLS. Consúltense los Protocolos 1.1–1.3 sobre el número de registros de inmuebles necesarios y el establecimiento de limitaciones en los rangos de fechas.
4. Completar las hojas de configuración: Regression Variables, Interactions, Aggregations, Calculations.
5. Ejecutar la regresión MARS (R/earth) para crear el modelo de precios.

6. Generar las contribuciones al valor de los inmuebles & los residuos (residuo puro y residuo/SF) a partir del modelo MARS.
7. Ordenar los inmuebles por residuo/SF.
8. Crear puntuaciones CQA para cada venta comparable.
9. Suponiendo que se ha producido una clasificación de calidad de los inmuebles que verdaderamente representa el atractivo general de los mismos, encontrar la mejor posición para el inmueble a valorar dentro de la clasificación y asignarle una puntuación CQA, estudiando las tendencias generales de los patrones en la clasificación.
10. Asignar el residuo/SF correspondiente al inmueble a valorar como residuo/SF · GLA (del inmueble a valorar).
11. **Sumar la estimación de precio de MARS y el residuo asignado para llegar a la conclusión de valor del inmueble a valorar.**
12. Descomponer el residuo del inmueble a valorar en componentes descriptivos y contribuciones al valor.
13. Calcular las contribuciones al valor & los ajustes producidos por el modelo para los comparables, además de todos los subtotales y los precios de venta ajustados.
14. Los precios de venta ajustados serán iguales a la conclusión de valor final del inmueble a valorar.
15. Seleccionar los 6-12 inmuebles más similares para el Sales Grid.
16. Descomponer los montos de residuo total de los comparables en contribuciones al valor de los componentes residuales, fijando el valor total del residuo en 0 una vez completada la descomposición.
17. Recalcular todos los subtotales y el precio de venta ajustado de los comparables del Sales Grid. Verificar los valores.
18. Cargar el Sales Grid al informe.

Obsérvese que, con el RCA, la conclusión de valor final NO se deriva directamente de los valores de los comparables en el Sales Grid. Con el método RCA, el Sales Grid se convierte en una herramienta destinada únicamente a explicar por qué inmuebles que parecen similares al inmueble a valorar se venden por más o por menos que la conclusión de valor final o el valor de avalúo del inmueble a valorar. Esto debe verse de otra manera: La conclusión de valor del inmueble a valorar depende de la estimación de precio del modelo MARS y del residuo estimado por el Valuation Engineer. Pero estas dos cantidades dependen en gran medida de todos los inmuebles que entraron en la regresión de R/earth, la cual creó la clasificación de residuos de los inmuebles que determinó la puntuación CQA del inmueble a valorar y, por consiguiente, determinó su residuo mediante la ubicación en la clasificación efectuada por el Valuation Engineer. Los ajustes en el Sales Grid están determinados, en realidad, por las restricciones matemáticas impuestas por el RCA; aunque la descomposición del residuo queda en cierta medida a merced del VE, los totales de las contribuciones al valor de la descomposición deben sumar siempre el residuo asociado. Sabemos por la Demostración 2 que, si bien los valores efectivos de los componentes residuales pueden variar según el juicio del VE, sus totales para un inmueble dado están completamente restringidos y, por lo tanto, no cambiarán el precio de venta ajustado del inmueble asociado, sin importar cómo altere el VE la descomposición del residuo.

5.3. Workflow de datos

Obsérvese que en este artículo se presenta únicamente un flujo de trabajo de alto nivel para el método RCA. Otros artículos profundizarán en las diversas etapas y subetapas del análisis RCA.

5.3.1. Requisitos de datos para MARS

Será necesario especificar una fecha inicial para las transacciones de venta y una fecha final basada en la fecha efectiva del avalúo o la valuación. Esto normalmente toma en consideración el número estimado de variables de regresión e interacciones que se necesitarán para cada grado de interacción superior a 1.

5.3.2. Protocolos: listings de ventas cerradas

Protocolo 1.1: Determinación del número requerido de listings de ventas cerradas del MLS

Este protocolo describe cómo calcular el número mínimo recomendado de listings de ventas cerradas que deben ingresarse a la regresión MARS. Será necesario hacer una estimación aproximada para comenzar, ejecutar MARS en ejecuciones de primer, segundo y posiblemente tercer grado para generar el modelo inicial, y luego revisar el número de filas conforme a las reglas siguientes para determinar si se necesitan más filas. Este protocolo supone que las variables independientes son 100% independientes, lo cual rara vez ocurre en bienes raíces; por ejemplo, la GLA (superficie habitable bruta) y los conteos de habitaciones suelen estar altamente correlacionados, y es probable que la GLA también muestre correlación con el tamaño del lote. Esto lleva al cálculo de los “Grados de Libertad (Degrees of Freedom, DOF).” Cada variable verdaderamente independiente recibe un grado de libertad. Pero, cuando existen dependencias entre las llamadas variables independientes, un cálculo exacto de los DOF requiere un análisis que está más allá del alcance de este artículo. Simplemente téngase presente que, debido a la probable colinealidad entre variables, las estimaciones siguientes probablemente queden de altas.

1. Para los términos de primer orden (sin interacciones), la regla de ~ 15 filas por variable funciona bien
2. Por cada interacción de dos vías permitida, deben añadirse aproximadamente 15-30 inmuebles
3. Para las interacciones de orden superior (3+ variables), cada una debería contar con 30+ inmuebles.

Sin embargo, esta no es una regla rígida. El número real necesario puede depender de:

- La relación señal-ruido de los datos
- La complejidad de las relaciones que se intenta modelar
- La variabilidad de la variable de respuesta
- La calidad de los datos

Ejemplo:

Así, si se tiene:

- 10 variables principales
- 5 interacciones de dos vías permitidas
- 2 interacciones de tres vías permitidas

El número mínimo de ventas podría calcularse como:

$$\begin{aligned} N &= (10 \text{ variables} * 15 \text{ filas/variable}) + \\ &\quad (5 \text{ interacciones-de-dos-vías} * 22 \text{ filas/interacción-de-dos-vías}) + \\ &\quad (2 \text{ interacciones-de-tres-vías} * 30 \text{ filas/interacción-de-tres-vías}) \\ &= 320 \text{ filas} \end{aligned}$$

Como alternativa, puede procederse de la siguiente manera: Después de ejecutar R/earth sobre un conjunto de datos, puede usarse el número de funciones base del modelo, más uno, como estimación de los grados de libertad, y luego multiplicar este número por 20-30 para obtener el número de registros de entrada necesarios.

Protocolo 1.2: Los períodos de venta deben abarcar cambios mínimos en las preferencias de los compradores

Al recopilar datos históricos de inmuebles, no existe una limitación temporal estricta si los cambios en las preferencias del mercado pueden capturarse mediante variables apropiadas. Considérense estos puntos clave para manejar la evolución arquitectónica en el mercado: Los cambios en los estilos arquitectónicos dominantes (como un desplazamiento del estilo California Ranch al Mediterráneo) pueden modelarse mediante:

- Identificadores de subdivisión
- Códigos de área del MLS
- Rangos de fecha de construcción

Sin embargo, si estos cambios de estilo no pueden capturarse adecuadamente mediante variables específicas, se enfrenta una restricción importante: si bien la fecha de venta se utiliza típicamente para ajustar por condiciones de mercado, emplearla simultáneamente para controlar las preferencias arquitectónicas confundiría estos dos efectos distintos. En tales casos, el enfoque preferido consiste en:

1. Limitar la recopilación de datos al período posterior a la introducción del estilo arquitectónico más reciente
2. Restringir la selección de comparables a subdivisiones que compartan el mismo estilo arquitectónico

Esto garantiza que el análisis mantenga la consistencia en las características de los inmuebles, permitiendo a la vez un ajuste adecuado por las condiciones de mercado a lo largo del tiempo. Como alternativa, al menos en este caso, suponiendo que el estilo arquitectónico pueda extraerse de los datos, podría considerarse la interacción entre estas dos variables. Sin embargo, tal extracción no siempre es posible.

Protocolo 1.3: En una zona con pocas ventas, haga lo mejor que pueda

Al trabajar con modelado MARS sobre datos de ventas, se necesitan aproximadamente 15 transacciones para un análisis significativo. Si se dispone de menos ventas (por ejemplo, 5 inmuebles), puede duplicarse cada registro dos veces para alcanzar el umbral de 15 inmuebles. Consideraciones importantes para este enfoque:

1. Utilizar únicamente regresión de primer grado para reducir el sobreajuste
2. Ignorar u omitir la validación cruzada, ya que el conjunto de datos es demasiado limitado
3. Si bien MARS probablemente superará a los métodos de pares emparejados (matched pairs) o de regresión lineal estándar en este escenario, téngase presente que la duplicación de datos afecta la validez de las mediciones de incertidumbre y de la significancia estadística

La mejor solución a largo plazo es ampliar el conjunto de datos con ventas reales adicionales. Sin embargo, si debe procederse con datos limitados, la duplicación de datos ofrece una solución temporal viable, reconociendo a la vez sus limitaciones estadísticas.

5.4. Estructura de carpetas del proyecto

Conviene analizar la estructura de carpetas del proyecto. Es una buena idea tener una carpeta separada para los análisis de mercado y otra para los avalúos de inmuebles individuales. Ambas carpetas tienen una estructura de subcarpetas similar:

```

RCA_Projects/
├── MarketAnalysis/
│   ├── Burlingame_20240501/
│   ├── Pacifica_20240310/
│   ├── RedwoodCity_20240615/
│   └── ...
├── SingleFamily/
│   ├── Pacific_Rosewood_123_20240510/
│   │   ├── Code/
│   │   │   ├── R/
│   │   │   │   └── ...
│   │   │   └── Python/
│   │   │       └── ...
│   │   └── Data/
│   │       ├── MLS_Original.xlsx
│   │       ├── MLS/
│   │       │   └── VER202402510153248/
│   │       │       ├── MLS.xlsx
│   │       │       ├── MLS_Stage2.xlsx
│   │       │       ├── MLS_Stage3.xlsx
│   │       │       ├── MLS_Stage4.xlsx
│   │       │       ├── MLS_Stage5.xlsx
│   │       │       └── Documents/
│   │       │           ├── LinearModels.pdf
│   │       │           ├── Cum_Dist.pdf
│   │       │           ├── Model.txt
│   │       │           ├── PartContrib.pdf
│   │       │           ├── Pairs.pdf
│   │       │           ├── Residual1.png
│   │       │           ├── Residual2.png
│   │       │           ├── ResidualSF.png
│   │       │           ├── Rpart.pdf
│   │       │           ├── VariablesImp.pdf
│   │       │           ├── VariableVsSP.pdf
│   │       │           └── ...

```

El procesamiento de la Etapa 1 creará la carpeta de versión con sus subcarpetas. Copiará el archivo `MLS_Original.xlsx`, aun cuando esta hoja de cálculo *no debería* cambiar nunca, sino reflejar exactamente lo que se descargó del MLS en una fecha y hora determinadas; no obstante, en un proyecto más prolongado existe la posibilidad de que se necesite una nueva descarga del MLS. A continuación, esta etapa hará una copia del libro `MLS.xlsx` expandido, con sus parámetros de configuración, dentro de la carpeta de versión. Puede haber muchas carpetas de versión, cada una creada para una ejecución completa de MARS.

Luego, las etapas siguientes generarán adiciones de columnas, ordenamientos y otros cambios

en el libro de Excel, y escribirán una versión actualizada como `MLS_Stage(N)` en la carpeta de datos del proyecto cuando el avalúo se haya completado. Para reiterar: el `MLS.xlsx` raíz experimenta cambios constantes, de modo que la copia de versión es una mutación continua del libro `MLS` original, con eliminaciones y adiciones de variables, además de otros cambios. El código de procesamiento en R o Python muy probablemente se ejecutará con la entrada del libro `Data MLS.xlsx`. De este modo, siempre se cuenta con una copia de todos los datos que respaldan cada versión de la ejecución del RCA.

Se crean tantas versiones como sea necesario hasta que ya no sea posible lograr mejoras, momento en el cual se selecciona la mejor versión como versión final. Podría cambiarse el nombre de la carpeta añadiendo “_Final” al final, de modo que resulte evidente al examinar el directorio. Algunas versiones anteriores con modelos interesantes o competitivos deberían normalmente conservarse. Cada carpeta de versión contiene una serie de archivos de texto, diagramas y gráficos que pueden incorporarse a un informe. El libro final `MLS_Stage5.xlsx` contiene tantas hojas como sean necesarias para el informe. En particular, puede tener una hoja de cálculo para transferirla a un Sales Grid de Fannie Mae. También contendrá necesariamente los cálculos o la descomposición detrás de las agregaciones. Esta es potencialmente la parte que más tiempo consume del enfoque RCA: la lucha por encontrar el modelo perfecto en una zona de mercado difícil.

Todo lo necesario para recrear los sales grids debe copiarse a la carpeta de versión, incluido el código R o Python utilizado — si el código experimenta cambios frecuentes.

5.4.1. Protocolos: múltiples ejecuciones de MARS (earth)

Protocolo 2.1: Ejecuciones con sobreajuste para estudiar quirúrgicamente los factores de precio de la zona de mercado

Durante las ejecuciones iniciales de earth, el sobreajuste debe usarse estratégicamente para descubrir los factores de precio subyacentes. Por ejemplo, una zona de mercado representada por 800 ventas puede contener solo 3 inmuebles con talleres, y en esos casos aislados el taller puede tener un impacto importante en el precio de venta final. Una ejecución de solo pasada hacia adelante (forward pass) producirá un modelo sobreajustado que muestra la contribución al valor aproximada de estos talleres. Sin embargo, dado que solo hay 3 ventas de este tipo entre 800, los términos de taller serán eliminados en la pasada hacia atrás (backward pass), y con certeza bajo validación cruzada. El valor que la pasada solo hacia adelante asigna a los talleres puede anotarse y transferirse al análisis de variables latentes, para usarse con cualquier inmueble con taller que aparezca en el Sales Grid. Estos modelos sobreajustados también proporcionan estimaciones detalladas de muchas otras características con datos escasos que serían eliminadas por la poda de la pasada hacia atrás o por las ejecuciones con validación cruzada.

Guárdese cada una de estas ejecuciones exploratorias como su propia carpeta de versión: las contribuciones al valor registradas son evidencia que el VE querrá citar más adelante al descomponer los residuos. Entiéndase, por supuesto, que el VE debe apoyarse en última instancia en la poda y la validación cruzada para lograr que el modelo final sea robusto frente a inmuebles que la regresión nunca vio — como un inmueble a valorar que aparezca posteriormente. Las ejecuciones sobreajustadas son instrumentos de estudio; nunca constituyen el modelo final.

Protocolo 2.2: Múltiples ejecuciones con validación cruzada bajo diferentes parámetros de earth()

Las zonas de mercado difieren en estructura y tamaño, y el VE normalmente necesita experimentar con una serie de ejecuciones: para desarrollar una comprensión de cuáles variables impactan más el precio; para juzgar si el conjunto existente de variables medidas cumple adecuadamente con la estimación del valor, o si deben añadirse nuevos datos (distancia al agua, elevación, distancia a centros comerciales, etcétera); y para descubrir nuevos tipos de interacciones, a la vez que se descartan otras por ineficientes o engañosas. Al final, el VE busca un buen R2 sobre los datos de entrenamiento y el mejor CVR2 posible frente a los datos de prueba. Llegar a ello puede requerir recopilar nuevos datos, mejorar la exactitud de los datos existentes o afinar los parámetros de earth() según el tamaño y el carácter de la zona de mercado (p. ej., según los grados de libertad disponibles — véase el Protocolo 1.1).

Precisamente por esto la estructura de proyecto anterior admite muchas carpetas de versión de ejecución: cada modelo candidato se conserva con sus parámetros y su R2/CVR2, de modo que la trayectoria a lo largo de las ejecuciones pueda compararse. Una versión cuyo R2 sube mientras su CVR2 se estanca o cae está ajustando ruido en lugar de señal de mercado, y el historial de versiones es lo que hace esto visible.

Protocolo 2.3: Expansión mediante Prolog de las columnas de comentarios y de características descriptivas

Aunque esta facilidad se encuentra todavía en una fase un tanto experimental, earthUI incorpora ahora un motor Prolog integrado (vProlog) y utiliza gramáticas de cláusulas definidas (Definite Clause Grammars, DCGs) para analizar sintácticamente las columnas de texto libre del MLS — Public Remarks ante todo, pero también los comentarios de los agentes y las listas de características delimitadas por comas, como las características de las habitaciones o las estructuras existentes. La expansión convierte frases como “owned solar,” “3 car garage,” o “40x50 workshop” en columnas de características ordinarias que pueden ofrecerse a la regresión, y las características con datos escasos recuperadas de esta manera alimentan directamente el análisis de variables latentes (véase la subsección posterior sobre las calificaciones de estado y calidad como regresores aproximados; la funcionalidad en sí se documenta en el artículo complementario sobre earthUI en este mismo número).

La expansión de texto multiplica el conjunto de predictores candidatos, y ello hace necesarias múltiples ejecuciones de earth/RCA para pruebas: las columnas expandidas deben probarse dentro y fuera del modelo, consolidarse en conteos o booleanos derivados cuando sean colineales, y verificarse contra el CVR2 para confirmar que portan señal de mercado y no ruido. Trátese cada uno de estos experimentos como su propia carpeta de versión de ejecución, exactamente como en los Protocolos 2.1 y 2.2. Y dado que la propia gramática de análisis sintáctico está en desarrollo, el VE debe revisar las columnas extraídas contra el texto fuente antes de confiar en ellas dentro de un modelo.

5.5. El libro de Excel principal

Dado que la mayoría de los servicios de MLS descargan los datos de los inmuebles a hojas de cálculo CVS de Excel, Excel es probablemente la forma más conveniente de almacenar los datos y los diversos parámetros de configuración. Existe un libro de Excel principal para cada proyecto, llamado `MLS.xlsx`, que tiene una hoja llamada “`MLSData`”, para almacenar los listings de ventas cerradas de inmuebles que se descargaron del MLS, así como varias hojas de configuración y parámetros:

```
Libro: MLS.xlsx/  
├─ Hoja: Project Data  
├─ Hoja: MLSData  
├─ Hoja: Regression Variables  
├─ Hoja: Calculations  
├─ Hoja: Allowed Interactions  
├─ Hoja: Aggregations  
└─ Hoja: (parámetros de ejecución de MARS)
```

Los datos originales se almacenan en un libro con un nombre similar a `MLS_Original.xlsx`. Debe hacerse una copia con un nombre diferente, digamos “`MLS.xlsx`,” que se cargará en el programa RCA de R o Python. Ambos libros se almacenan bajo la carpeta `Data`. Ahora bien, el libro `Data/MLS.xlsx` probablemente pasará por muchas iteraciones de ejecuciones, probando distintas combinaciones de variables, parámetros de MARS y cambios en otras configuraciones. Cada iteración crea una nueva carpeta de entrada/salida nombrada con la fecha y la hora, con el formato `VER_YYYYMMDDHHMMSS`. Cada carpeta tendrá una copia del libro `MLS.xlsx` de entrada y contendrá los libros de Excel generados para cada una de las cuatro etapas de procesamiento (Etapas 2–5), así como los documentos de texto, diagramas y gráficos de salida. Obsérvese que la Etapa 5, la etapa final, puede contener varias hojas adicionales para el informe, según el tipo de cliente y el SOW (alcance del trabajo).

5.6. Copias de seguridad

Al escribir hojas de cálculo de Excel nuevas o modificadas en las carpetas de versión del proyecto, también pueden hacerse copias de seguridad en bases de datos SQLite o PostgreSQL dentro de la carpeta del proyecto. Los archivos de base de datos tienen menos probabilidades de corromperse involuntariamente en períodos posteriores, durante su uso para revisión o consulta.

6. Notación del libro de Excel

6.1. Dimensiones descriptivas

Existen aproximadamente seis dimensiones descriptivas para un libro de Excel “M”:

1. Avalúos “a”
2. Versiones de ejecución “r”
3. Etapas del flujo de trabajo “w”
4. Hojas “s”
5. Listing de propiedades “p”
6. Conjunto de variables “v”

Así, una celda individual de un libro de Excel particular M puede describirse como:

$$\begin{matrix} r_* \\ w_* \end{matrix} \begin{matrix} a_* \\ W \\ s_* \end{matrix} \begin{matrix} p_* \\ v_* \end{matrix} \quad (1)$$

(Nota: “*” significa algún índice válido)

o bien:

$$\begin{matrix} r_k \\ w_m \end{matrix} \begin{matrix} a_g \\ W \\ s_h \end{matrix} \begin{matrix} p_i \\ v_j \end{matrix} \quad \begin{matrix} i = 1, \dots, n_p \\ j = 1, \dots, n_v \\ k = 1, \dots, n_r \\ m = 1, \dots, n_m \\ g = 1, \dots, n_g \\ h = 1, \dots, n_h \end{matrix} \quad (2)$$

donde:

- n_p = # de filas (depende de la hoja)
- n_v = # de columnas (depende de la hoja)
- n_r = # de versiones de ejecución
- n_m = # de etapas del flujo de trabajo
- n_g = # de proyectos de análisis
- n_h = # de hojas del libro de Excel

Entiéndase que “W” hace referencia al libro de Excel completo, compuesto por seis hojas de datos centrales, cada una con su propia tabla, las cuales, al ejecutarse el programa asociado en R o Python, se cargan en data frames internos separados. Tanto R como Python cuentan con funciones internas capaces de tomar un nombre y determinar a qué fila o columna del data frame pertenece. Por ejemplo, si en el subíndice v_j , $v = (\text{“GLA”, “LotSize”, “BathRms”, “BedRms”, ...})$ y $j=0$, entonces $v_j = \text{“GLA”}$. Tanto R como Python podrán resolver internamente “GLA” hacia la columna en la que está almacenada, que, por ejemplo, podría ser la columna 15 del data frame asociado. Así, en este artículo indexamos nuestros arreglos de nombres, los cuales pasan un nombre al programa en ejecución, que reindexa dicho nombre hacia el índice real de columna o fila del data frame interno. Es un tanto complicado. Pero, para ser precisos, esto es necesario y sumamente conveniente. Se trata de una “separación de responsabilidades” entre el rol del analista y el rol del desarrollador; aunque en la práctica el Valuation Engineer (VE) podría desempeñar ambos roles. Nótese que un “data frame” de R se denomina “dataframe” en Python.

Nótese también que la hoja de parámetros de ejecución de MARS no constituye una dimensión de W, ya que W se relaciona con la versión de los datos que se introducen en MARS, independientemente de los parámetros que determinan cómo MARS/earth procesa dichos datos.

En aras de la claridad, se presenta una notación multidimensional compleja para el libro de Excel, la cual se relaciona de manera más directa con los data frames internos de R o Python en los que se cargan las hojas de datos del libro. Es una notación “pesada”. Pero cuando lleguemos a las demostraciones y demás, descartaremos gran parte de ese lastre mediante supuestos sobre aquello con lo que estamos trabajando. No obstante, esta notación ayuda a obtener desde el inicio una mejor visión general del flujo de trabajo y de los datos.

Comencemos con la notación para el conjunto de libros de Excel que procesaremos.

6.1.1. La dimensión de proyecto (“a”)

Una vez aprobado el alcance del trabajo (SOW), es necesario configurar una carpeta de proyecto para almacenar los datos de entrada y salida, incluidos los parámetros de configuración, los gráficos de salida, los diagramas, las hojas de cálculo y los fragmentos de informe. El VE debe identificar todos los proyectos con un ID secuencial simple, como un número entero, de modo que si un informe llegara a extraviarse, resultaría evidente. Si un proyecto se cancela, debe existir un medio consistente para indicar dentro de la carpeta del proyecto que este ha sido cancelado, pero dicha carpeta nunca debe eliminarse. Suponemos que los ID de proyecto comienzan en 0 y se incrementan de 1 en 1, aunque existen otras posibilidades. Suponemos además que todos los ID de proyecto válidos se almacenan en un conjunto ordenado “a” creado con el programa de procesamiento.

Los datos del proyecto de valuación se encuentran en la hoja de cálculo de proyecto del libro de Excel, pero también podrían almacenarse en alguna base de datos SQL central,

A continuación se sugieren los siguientes campos de proyecto:

1. ID local: Un ID entero corto que se genera secuencialmente para la empresa. Constituiría una excelente clave primaria para el almacenamiento - y podría ser generado automáticamente por la base de datos con cada proyecto sucesivo.
2. GUID global: Si se desea, puede crearse un GUID (Globally Universal ID) mediante

<https://www.uuidgenerator.net/>

3. Ubicación del inmueble (análisis de mercado) o dirección
4. Fecha efectiva del avalúo
5. Fecha del encargo
6. Fecha de inspección
7. Fecha del informe original
8. Fecha del primer informe actualizado
9. Fecha del segundo informe actualizado
10. Fecha del tercer informe actualizado
11. Nombre del cliente
12. Dirección del cliente
13. Nombre del prestamista
14. Dirección del prestamista
15. Nombre del propietario
16. Dirección del propietario
17.

Ejemplo de lista de ID de proyecto:

```
a = ("1",
      "2",
      "...",
      "141",
      "142")
```

6.1.2. Dimensión de versión de ejecución (“r”)

En segundo lugar dentro del flujo de trabajo, ejecutamos la regresión MARS sobre los datos de entrada. Esto probablemente implicará desde varias hasta muchas ejecuciones, a medida que modificamos diversos parámetros para mejorar el modelo. Cada ejecución creará una nueva carpeta de versión para la salida, con un nombre basado en la fecha y hora de ejecución. Por ejemplo,

```
r = ("ver20241210_083124",
      "ver20241209_151130",
      "ver20241208_160944")
```

```
latest_version : r1 = "ver20241210_083124"
```

Las versiones prometedoras, con modelos y salidas competitivos, deben conservarse hasta que se decida una versión final, o quizá más tiempo si tienen alguna utilidad. Puede resultar conveniente modificar la carpeta de versión para indicar que se trata de la versión final, colocando “Final” al final del nombre.

6.1.3. Etapas del flujo de trabajo (“w”)

El flujo de trabajo del RCA se divide habitualmente en cinco etapas separadas, según se enumeran a continuación. Cada etapa debe completarse con éxito antes de iniciar la siguiente. El VE debe poder detener el procesamiento al final de una etapa para revisar los resultados antes de continuar. Esto es necesario porque las zonas de mercado son bastante complejas y, con frecuencia, muy singulares. Pueden presentarse problemas inusuales que exijan un enfoque nuevo. El VE debe poder intervenir en el código de R o Python para realizar los cambios necesarios e incluso probar nuevos enfoques. Y, en tales casos, es posible que desee experimentar y probar diferentes enfoques que requieran múltiples ejecuciones de la misma etapa en la que está trabajando. No obstante, el VE también debe tener la libertad de ejecutar todas las etapas con un solo comando, - con algunas salvedades (p. ej., la agregación de tercer grado debe haberse determinado previamente mediante una ejecución anterior de la Etapa 2). Esto significa que, al completarse cada etapa, esta debe almacenar todos sus datos internos en archivos o bases de datos, de modo que la siguiente etapa pueda iniciarse como si la etapa anterior aún estuviera en memoria, recargando en memoria los datos de la etapa previa.

Estas son las etapas:

```
w = ("stage1", -Configuración
      "stage2", -Procesamiento MARS
      "stage3", -Sales Grid final
      "stage4", -Fragmentos de informe
      "stage5") -Revisión/Limpieza
```


stage2 : $w_2 = \text{"stage2"}$

6.1.4. Hojas del libro de Excel (“s”)

En el libro de Excel principal (Core) del RCA, contamos con las siguientes hojas, cada una de las cuales se mapea a un data frame interno:

```
s = ("MLSDData",
      "Regression Variables",
      "Calculations",
      "Interactions",
      "Aggregations 1st & 3rd Degree",
      "Aggregations 2nd Degree")
```

La razón por la que las agregaciones de segundo grado se colocan en una hoja separada es que pueden definirse con mayor facilidad mediante una tabla simétrica que muestra todos los pares posibles de interacciones, lo que permite especificar, para cada par, una variable de agregación a la cual se agregan (suman) las contribuciones al valor y los ajustes asociados. No podríamos hacer esto con las agregaciones de tercer grado, porque ello requeriría una tabla simétrica tridimensional, lo cual no es posible en Excel. Además, una tabla de 20x20x20 supondría manejar 8,000 celdas. Resulta más sencillo simplemente listar los nombres de los términos de tercer grado en la tabla junto con las agregaciones de primer grado, por ejemplo:

Variable	Agregar a
AboveGradeFinArea	GLA
BelowGradeFinArea	BGLA
UnfinArea	UBA
LotSize	LotSize
...	...
GLA__LotSize	GLA
GLA__DateOfSale	GLA
BathRms__DateOfSale	GLA
Lat__Long	Location
Lat__Long__GLA	Location
...	...

6.1.5. Listings de propiedades (“p”) - filas de MLSDData

Los listings de propiedades constituyen las filas de la hoja MLSDData. No se encuentran en las demás hojas hasta la Etapa 4, en la que se crean los fragmentos finales del informe con la selección final de inmuebles comparables. A continuación se muestra una lista de ejemplo de ID de propiedades:

```
p = ("MLS823211", "MLS823456", "MLS823542", "MLS823721", "...")
```

6.1.6. Variables MLS (“v”) Columnas de MLSDData

Más adelante se dirá más sobre el manejo de variables. Baste señalar por ahora que la hoja MLSDData contiene típicamente entre 30 y 40 variables de diversos tipos, aunque por lo general solo se seleccionarán de 10 a 20 como variables independientes para la regresión MARS. Estas variables también aparecen en otras hojas, aunque típicamente como una lista de todas las variables posibles que pueden utilizarse para su envío a MARS. Un ejemplo de lista de variables es:

$$v = (\text{“GLA”}, \text{“SalePrice”}, \text{“BedRms”}, \text{“BathRms”}, \text{“GarageSF”}, \text{“...”})$$

6.1.7. Ejemplo final de notación

Reunamos la notación anterior en un ejemplo final. Utilizando las listas anteriores, sea

$${}_{w_1}^{r_1} W_{s_0}^{a_{141}} {}_{v_0}^{p_2} \quad (3)$$

Nótese que aquí utilizamos indexación desde 0, de modo que se supone que el primer elemento de las listas tiene índice 0. Python también utiliza indexación desde 0, pero R utiliza indexación desde 1.

Así, lo anterior significa que nos referimos al valor de la celda correspondiente a:

- La variable v_0 o columna “GLA”
- Para la propiedad p_2 o ID “MLS823542”
- De la hoja s_0 o “MLSDData”
- Para la etapa del flujo de trabajo w_1 o “stage2”
- De la versión r_1 o
“ver20241209_151130”
- Del proyecto a_{141} o “142”

La notación anterior puede entonces escribirse también como:

$${}_{\text{“stage2”}}^{\text{“ver20241209_151130”}} W_{\text{“MLSDData”}}^{\text{“142”}} {}_{\text{“GLA”}}^{\text{“MLS823542”}} \quad (4)$$

6.1.8. Justificación

“¿Por qué no usamos simplemente índices o números en la notación?” Tenemos dos requisitos para nuestra notación:

1. Subconjuntos; p. ej., el analista selecciona con frecuencia subconjuntos de una lista, tales como subconjuntos de variables. Por ejemplo, puede haber 35-40 variables en la hoja MLSDData, pero solo se usan 10-20 de ellas en una ejecución determinada. De una ejecución a la siguiente, el VE cambia con frecuencia las variables que se someten a MARS/earth.

2. Iteración; p. ej., $i = 1, 2, \dots, n_1$

Nuestros algoritmos, protocolos y demostraciones normalmente necesitan poder iterar sobre todos los elementos de una lista ordenada, del primero al último — o bien seleccionar algún elemento en particular.

Así pues, necesitamos una letra que represente el subconjunto o la lista de elementos, y esta debe llevar otro subíndice que podamos usar para iterar sobre los elementos de la lista, de modo que nos proporcione las cadenas de texto que se utilizan para indexar las filas o columnas del data frame.

6.1.9. ¿Matrices?

¿Y qué hay de las matrices del álgebra lineal? Las matrices funcionan únicamente con números, ya sean enteros o reales. Las matrices como estructura de datos están soportadas tanto en R como en Python. Sin embargo, muchas de las variables con las que trabajamos en los datos de la MLS son lo que se denomina “strings” o datos de tipo carácter. Incluso si algunas variables son solo números, a menudo no son números reales sino simplemente nombres. Por ejemplo, los códigos de área de la MLS para una ciudad pueden ser (“660”, “661”, “662”, “663”, $\dots$, “672”). La regresión de earth permite especificar variables como variables “factor”, en cuyo caso los números se tratan simplemente como nombres, sin secuencia alguna. En tales casos, el modelo producido se limitará a proporcionar una contribución al valor para cada número, en lugar de una función que recorra la secuencia. Otros ejemplos incluyen “C1”, $\dots$,”C6” para el estado (Condition) y “Q1”, $\dots$,”Q6” para la calidad (Quality).

6.1.10. Primera simplificación

La notación anterior para el libro de Excel central del proyecto resulta útil para vincular los elementos entre sí, pero es demasiado pesada para llevarla consigo al explicar algoritmos o demostraciones matemáticas. Cuando realizamos el trabajo tedioso, podemos suponer que este se desarrolla dentro del proyecto, la versión, la etapa y la hoja dados. Simplemente indicamos con qué estamos trabajando, si no está ya implícito. Todo lo que realmente necesitamos de lo anterior, una vez que indicamos con qué estamos trabajando, es:

$$M_{s_h}^{p_i v_j} \quad \begin{array}{l} i = 1, 2, \dots, n_p \\ j = 1, 2, \dots, n_v \\ h = 1, 2, \dots, n_s \end{array} \quad (5)$$

donde: n_p = # de propiedades
 n_v = # de variables de entrada
 n_s = # de hojas de entrada

6.1.11. Segunda simplificación

Resulta más conveniente renombrar las hojas según los tipos específicos de data frames que soportan, en lugar de usar índices. En cada hoja, las filas y las columnas tienen funcionalidades distintas. Así, el conjunto W de data frames se descompone en los data frames M, V, A, I y C de la siguiente manera:

$$\diamond \text{ Datos MLS} \quad M_{v_j}^{p_i} = W_0^{p_i v_j} \quad \begin{array}{l} i = 0, 1, \dots, n_p \\ j = 0, 1, \dots, n_v \end{array} \quad (6)$$

donde: n_p = # de propiedades
 n_v = # de variables de entrada

Nota: El inmueble a valorar es siempre, por defecto, el primero en las listas de propiedades

En la hoja de cálculo de Excel “MLSDData” y su data frame asociado, la primera fila está siempre reservada para el inmueble a valorar. Nunca se ordena junto con las demás propiedades de la MLS que aparecen debajo. Nunca se introduce en la regresión MARS como dato independiente; se introduce más bien como dato objetivo, o como la variable dependiente. Lo mismo se aplica a las listas de propiedades. En todas las listas de propiedades se asume siempre que el primer elemento, con índice 0, es el ID del inmueble a valorar, salvo que se indique lo contrario.

Incluso si la tarea consiste en crear un modelo para una zona de mercado sin ningún inmueble a valorar, la primera fila se carga de todos modos con datos ficticios de alguna propiedad. Esto hace mucho más conveniente escribir programas, especificar algoritmos o realizar demostraciones. Así, si usamos c para representar la lista de propiedades, c_0 es el inmueble a valorar y $c_i, i = 1, \dots, n_p$ son los listings comparables que se introducirán en la regresión MARS/earth.

◇ **Especificaciones de variables de regresión**

$$V_{s_j}^{v_i} = W_1^{v_i}_{s_j} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_s \end{matrix} \quad (7)$$

donde: n_v = # de variables disponibles
 n_s = # de especificaciones

Obsérvese que $n_s \leq n_v$, ya que normalmente seleccionamos un subconjunto de las variables disponibles como entrada a la regresión.

◇ **Cálculos**

$$C_{c_j}^{v_i} = W_2^{v_i}_{c_j} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_c \end{matrix} \quad (8)$$

donde: n_v = # de variables
 n_c = # de especificaciones de cálculo

◇ **Interacciones**

$$I_{t_j}^{f_i} = W_3^{f_i}_{t_j} \quad \begin{matrix} i = 1, 2, \dots, n_f \\ j = 1, 2, \dots, n_t \end{matrix} \quad (9)$$

donde: n_f = # de variables de origen
 n_t = # de variables de destino

Obsérvese que el data frame I es un data frame simétrico, con las variables listadas alfabéticamente en las filas “from” (origen) y las columnas “to” (destino). Las celdas se marcan con un “1” para indicar que se permite una interacción entre el par de variables asociado. En caso contrario, un “0” indica que no se permite interacción alguna. Para una interacción de segundo grado, esto es relativamente simple, con una sola interacción posible. Con una regresión de tercer grado, tenemos tres interacciones por pares entre las tres variables, donde cada par de interacciones debe estar permitido para generar el término de interacción (o variable) de las tres variables.

◇ **Agregaciones**

$$A_{v_j}^{v_i} = W_4^{v_i}_{v_j} \quad \begin{matrix} i = 1, 2, \dots, n_v \\ j = 1, 2, \dots, n_v \end{matrix} \quad (10)$$

donde: n_v = # de variables

7. El Sales Grid del RCA

Téngase presente que el protocolo RCA requiere un programa en R o Python capaz de acceder a R/earth y ejecutar todos los pasos detallados. Es un proceso intensivo en datos y el trabajo no podría realizarse manualmente de ninguna manera. Al mismo tiempo, suponiendo que el programa esté libre de errores, se necesita un programa de software para garantizar que el trabajo requerido se realice con precisión y sin error humano.

También es una necesidad que los VE que realizan este trabajo sean capaces de modificar dichos programas según se requiera, especialmente cuando se enfrentan a zonas de mercado, vecindarios, inmuebles u otros requisitos o problemas nuevos e imprevistos.

Dentro del procesamiento del RCA tenemos tres tipos de Sales Grids:

1. El Sales Grid de Procesamiento del RCA, denominado en este artículo “MLSDData,” que es un data frame de gran tamaño, con su hoja de cálculo asociada, que puede contener numerosos detalles intermedios que finalmente pasan por un proceso de agregación para llegar al Sales Grid del RCA que se menciona más abajo. Un ejemplo sería demasiado grande para mostrarlo aquí. Lo más probable es que los Valuation Engineers adapten esta hoja de cálculo a sus preferencias. Tengo la intención de escribir en el futuro un artículo aparte sobre el desarrollo de un Sales Grid de este tipo.
2. El Sales Grid Final del RCA, que se utiliza en informes no destinados a las GSE y que se asemeja a la [Table 2](#). Este imita, por ejemplo, el formulario de Fannie Mae, lo suficiente como para ofrecer al lector cierta base de familiaridad. Sin embargo, obsérvese que contiene más columnas por comparable de venta que el cuadro de Fannie Mae: cada comparable tiene una columna de “Contribución al valor” junto a sus columnas de “Medida” y “Ajuste,” y el inmueble a valuar también cuenta con una — columnas de las que carece el formulario de Fannie Mae. Un formato más sencillo es también el de la [Table 1](#).
3. El Sales Grid de Carga para las GSE, que se utiliza para cargar el Sales Grid en software como Alamode. Nótese, sin embargo, que no todos los paquetes de software de formularios de avalúo ofrecen la posibilidad de cargar los datos del Sales Grid desde una hoja de cálculo. Por necesidad, dichos grids no pueden incluir una columna separada para las contribuciones al valor, de las cuales las GSE como Fannie Mae no tienen comprensión alguna. La estrategia consiste en entregarles el formulario que exigen, pero respaldarlo con un Sales Grid del RCA, así como con otra documentación de apoyo, tal como los detalles de cualquier agregación de variables.

Los Sales Grids de Procesamiento del RCA no se parecen a los Sales Grids de Fannie Mae, sino que son, aproximadamente, una transposición del Sales Grid de Fannie Mae, con las filas y columnas intercambiadas: cada fila contiene los datos de algún inmueble y cada columna contiene los datos de alguna variable del inmueble. Fácilmente puede haber cientos o miles de filas de inmuebles y 200 o más columnas. La primera fila se reserva para los nombres de las variables, y la segunda para un inmueble a valuar, incluso si solo se está realizando un análisis de mercado. Cuando la hoja de cálculo se lee en un data frame, la primera fila de la hoja se convierte en el atributo names, y la segunda fila, correspondiente al inmueble a valuar, se convierte en la primera fila del data frame.

Las columnas del Sales Grid del RCA incluyen columnas de identificación y, a continuación, las variables primarias, externas y sintéticas, combinadas e introducidas como variables independientes en MARS. A partir de estas variables, MARS produce como salida aquellas que considera que tienen un impacto significativo sobre el precio de venta, también conocido como variable dependiente u objetivo. En la regresión de segundo grado, podemos obtener pares de

estas variables que actúan como nuevas variables. Con la regresión de tercer grado, también podemos obtener tripletes de variables que actúan como supuestas variables independientes, aunque, por supuesto, están altamente correlacionadas con las variables subyacentes.

8. Variables

En el lado de entrada de la ejecución de MARS, contamos únicamente con variables simples, que son:

1. Características del inmueble cargadas desde la hoja de cálculo del MLS.
2. Variables derivadas creadas por el VE para refinar las variables del MLS o para añadir transformaciones de variables existentes. Por ejemplo, el VE puede preferir transformar la Date of Sale en “DaysOffMarket”, que es igual a la fecha efectiva del avalúo menos la Date of Sale.
3. El VE puede importar datos de fuentes de terceros, como las tasas de interés hipotecarias. O bien, si faltan datos de latitud, longitud o elevación en los datos del MLS, puede importarlos de otras fuentes. Estas variables podrían denominarse “variables externas”.

En el lado de salida de MARS, trabajamos con un modelo de precios que es una combinación de términos que:

1. Contienen solo una de las variables de entrada y proporcionan la ecuación de la contribución al valor de dicha variable. Estos se denominan términos (o variables de salida) de primer grado. El término mismo se trata como un valor de contribución al valor y se le asigna un nombre.
2. Contienen dos de las variables de entrada y constituyen la ecuación de la contribución al valor de la interacción de las dos variables dadas. Estos términos pueden denominarse términos (o variables de salida) de segundo grado.
3. Contienen tres de las variables de entrada y constituyen la ecuación de la contribución al valor de la interacción entre las tres variables. Estos términos pueden denominarse en este artículo términos (o variables de salida) de tercer grado.

Las variables de salida de MARS son, entonces, contribuciones al valor, incluida la contribución al valor residual. A su vez, se tratan como variables en los cálculos del Sales Grid, tales como el cálculo de ajustes o la creación de gráficos.

Podemos tener distintos tipos de variables que se encuadran en dos ontologías básicas: (1) el origen de la variable y (2) la aplicación de la variable, es decir, no residual frente a residual

8.1. Ontología de origen de las variables

1. Variables de primer grado
 - 1.1 Variables de regresión del MLS
 - 1.2 Variables de regresión externas
 - 1.3 Variables de regresión derivadas

2. Variables factor:

Para manejar las variables factor, se necesita una función capaz de identificarlas. Dado que se especifican en el libro de Excel de configuración del proyecto, se crea una función que devuelve verdadero si la variable que se le pasa ha sido clasificada como de tipo factor.

3. Variables de interacción: Estas son creadas por MARS al generar el modelo, si el grado se establece en 2 o superior. Sin embargo, 3 variables en interacción es lo máximo que podría recomendarse debido a la dificultad de interpretación, la probabilidad de sobreajuste y el impacto sobre los grados de libertad, con el consiguiente aumento de los registros de inmuebles necesarios para garantizar un modelo robusto.
 - 3.1 Variables de segundo grado: Estas son comunes y buenos ejemplos serían GLA:BathRms, GLA:LotSize, DateOfSale:BathRms, Latitude __Longitude.
 - 3.2 Variables de tercer grado: Interacciones menos comunes que sugieren cautela y posible sobreajuste.
4. Variables de descomposición del residuo

8.2. Ontología de aplicación de las variables

1. Contribuciones al valor generadas por MARS
 - 1.1 Variables de primer grado
 - 1.1.1. Variables de regresión del MLS
 - 1.1.2. Variables de regresión externas
 - 1.1.3. Variables de regresión derivadas
 - 1.2 Variables de interacción
 - 1.2.1. Variables de segundo grado
 - 1.2.2. Variables de tercer grado
 - 1.3 Variables de descomposición del residuo
 - 1.4 Variables de agregación
2. Variables & valores generados por el ingeniero
 - 2.1 Variables de descomposición del residuo
 - 2.1.1. Diseño/Estética
 - 2.1.2. Paisajismo
 - 2.1.3. Utilidad funcional
 - 2.1.4. Vista
 - 2.1.5. Estado
 - 2.1.6. Calidad
 - 2.1.7. ...

A estos grupos de variables se les asignará su notación y se describirán con mayor detalle en la siguiente sección

Protocolo 3.1: Variables factor

Las variables factor, o simplemente factores, pueden aparecer en todos los términos, incluidas las interacciones. Son importantes porque deben recibir un tratamiento especial en la regresión MARS y en la elaboración de gráficos.

Si bien las variables con valores de tipo carácter son necesariamente factores, las variables numéricas deben marcarse como variables factor para que MARS las reconozca como tales. Por ejemplo, las zonas del MLS suelen identificarse mediante un número. En particular, las zonas de vecindario en el San Francisco Bay Area se identifican a menudo con enteros únicos. Si estos valores se pasan a MARS como factores, MARS intentará descubrir si ciertas zonas tienen un valor contributivo promedio sobre las ventas de inmuebles dentro de sus límites, sin considerar relación alguna con otras zonas. En cambio, con las variables normales, asumirá que existe algún tipo de continuidad en las contribuciones al valor a medida que cambia el valor de la variable. Esta es una distinción importante. Además, estas variables factor pueden aparecer en términos del modelo junto con variables no factor, lo que dificulta la elaboración de gráficos. En concreto, al graficar términos de una sola variable con factores, agrupamos todos esos factores y simplemente dibujamos un gráfico de barras que muestra sus contribuciones, positivas o negativas.

Si los factores aparecen con otras variables en los términos, siempre aparecerán con un único valor posible, porque las variables factor solo pueden tener un valor para un inmueble dado: si una zona tiene el valor 662, por definición no puede tener ningún otro valor para esa zona. Si esto le resulta confuso, para complicar más las cosas, en el modelo generado por MARS no existe un único nombre de variable para el factor; MARS tomará el nombre raíz de la variable factor que recibió como entrada y luego creará nuevas columnas con nombres de variable formados al anexar los valores posibles. Así, una sola columna en la hoja de cálculo de Excel, como `MlsArea`, tendrá múltiples columnas del data frame creadas para cada valor posible, como `MlsArea660`, `MlsArea661`, etc. Una variable como `MlsArea661` tendrá el valor 1 para los inmuebles en la zona 661; de lo contrario, 0. Después, MARS efectuará la regresión sobre esas columnas como si representaran variables booleanas. Estos nombres modificados se utilizarán en el modelo proporcionado por MARS.

Probablemente querrá convertir esos nombres del modelo de vuelta a su nombre original para que los gráficos resulten más comprensibles para los clientes, a quienes les gusta lo claro y sencillo (al menos tan sencillo como sea posible). En lugar de ver `MlsArea661=1`, veremos `MlsArea=661`. Entonces podemos graficar los términos indicando simplemente en el encabezado del gráfico algo como “Para MLS Area = 662”. De este modo, los términos de grado N con M variables factor se grafican como términos de grado (N-M), y la dimensión del factor desaparece en el título.

Obsérvese que dos o más variables factor podrían existir en un mismo término, por lo que probablemente se necesitaría un gráfico para cada par posible de valores. Si la variable x, con tres valores distintos, y la variable y, con dos valores distintos, aparecen en el mismo término con la variable no factor z, entonces se necesitarían $2 \cdot 3 = 6$ gráficos. Un ejemplo sería “`MlsArea`” + “`Architectural Style`”; p. ej., las casas victorianas en San Francisco podrían valer más o menos según la zona de San Francisco en la que se encuentren. Afortunadamente, en la mayoría de las zonas suele haber solo 1-3 estilos arquitectónicos lo suficientemente importantes como contribuyentes al valor para ser utilizados en el modelo (aunque puede haber 6+ estilos arquitectónicos diferentes).

8.3. Términos de segundo y tercer grado

Con la regresión de segundo o tercer grado, algunos de los términos del modelo de salida pueden incluir dos o, respectivamente, tres variables. Por ejemplo:

$$\text{BathRms_GLA} = -52 * \max(0, 1200 - \text{GLA}) * \max(0, \text{BathRms} - 1)$$

Obsérvese que estos términos se nombran ordenando alfabéticamente los nombres de las variables y concatenándolos, quizás con un guion bajo. Hacemos lo mismo con los términos de tercer grado sobre las tripletas de nombres de variables producidas en el modelo. Por ejemplo, “BathRms_GLA_LotSz.”

Estas variables se añaden a las variables de entrada primarias en la Etapa 3, cuando el programa RCA calcula las contribuciones de todas las variables para cada inmueble, incluidas las variables de interacción.

8.3.1. Variables de ajuste

El programa RCA también calculará los ajustes para todas las variables e inmuebles en la Etapa 3, después de calcular las contribuciones al valor, restando la contribución al valor del inmueble comparable correspondiente a una característica de la contribución del inmueble a valuar para esa característica.

8.3.2. Variables de agregación

Las variables de agregación se definen para reemplazar subconjuntos de las variables, sumando todas sus contribuciones al valor y ajustes en una variable común que se utilizará en el informe. Por ejemplo, podríamos haber dividido tanto el área habitable terminada como la no terminada en variables de mayor detalle:

1. Área terminada legal sobre rasante
2. Área terminada legal bajo rasante
3. Área terminada ilegal sobre rasante
4. Área terminada ilegal bajo rasante
5. Área no terminada sobre rasante
6. Área no terminada bajo rasante

Sí, hay zonas donde esto tiene sentido.

La regresión MARS efectuará la regresión sobre todas las variables y generará un modelo que proporcionará contribuciones y ajustes para ellas. Pero para un informe en formulario, estas distintas contribuciones probablemente tendrán que combinarse o agregarse en menos variables. Los detalles de la agregación deben proporcionarse en un anexo. Un ejemplo más frecuente sería agregar los ajustes de interacción en uno de los ajustes de las variables primarias. Por ejemplo, podríamos agregar el ajuste de GLA_Lotsize a GLA o a LotSize, o incluso dividirlo 50:50 entre ambos, según lo determinado en la hoja de cálculo Aggregation.

8.3.3. Variables especiales & notación

π_i	Precio de venta neto real del inmueble $i = 1, 2, \dots, n_p$; para el inmueble a valuar, π_0 denota la conclusión de valor (véase la Demostración I)
α_i	Precio de venta neto ajustado del inmueble $i = 1, 2, \dots, n_p$
ρ_i	Precio de venta neto estimado por MARS para el inmueble $i = 0, 1, \dots, n_p$
β	Valor base del modelo
$\varphi_{i,j}$	Contribución al valor de la variable de regresión $j = 1, 2, \dots, n_r$ para el inmueble $i = 0, 1, \dots, n_p$
$\delta_{i,j}$	Ajuste de la variable de regresión $j = 1, 2, \dots, n_r$ para el inmueble $i = 1, 2, \dots, n_p$
ϵ_i	Residuo total del inmueble $i = 0, 1, \dots, n_p$
$\epsilon_{i,c}$	Contribución al valor del componente residual c , $c = 1, 2, \dots, n_c$ para el inmueble $i = 0, 1, \dots, n_p$
$\xi_{i,c}$	Ajuste del componente residual c , $c = 1, 2, \dots, n_c$ para el inmueble $i = 0, 1, \dots, n_p$

donde

$n_p = \#$ of comparable properties

$n_r = \#$ of regression variables

$n_c = \#$ of residual breakdown variables

8.3.4. Descomposición de los residuos

Una vez que la regresión MARS ha creado un modelo, este genera los precios de venta estimados a partir de las variables de entrada seleccionadas por MARS. Resta ese valor del Net Sale Price correspondiente para obtener un valor residual para cada inmueble. Ahora bien, el protocolo RCA estipula que, una vez que el VE ha decidido los comparables definitivos, por lo general 6-12 listings de inmuebles, debe revisar cada listing y descomponer el residuo calculado en las características más probables (variables latentes) que, a su juicio, impactan el valor, ignorando la posibilidad de errores de datos aleatorios. Podría, por ejemplo, tener estas variables de

descomposición para un inmueble dado:

1. Diseño
2. Utilidad funcional
3. Venta por encima/por debajo del mercado
4. Estado
5. Calidad

Es probable que distintos inmuebles tengan algunas variables de descomposición del residuo diferentes, las cuales deben reunirse en un único conjunto ordenado.

Estimación de las contribuciones de características infrecuentes. Algunas características portadoras de valor son demasiado infrecuentes para el canal de regresión: un taller presente en 3 de 780 ventas no puede obtener un coeficiente defendible, y el pasadizo hacia adelante normalmente omitirá por completo dicha variable. Estas son características residuales, y sus contribuciones pueden estimarse *en el espacio de los residuos*: se ajusta el modelo sin la característica y luego se efectúa la regresión de los residuos del modelo sobre la(s) columna(s) excluida(s) mediante mínimos cuadrados ordinarios. Para un indicador binario, el coeficiente es una contribución de suma global (equivalentemente, el desplazamiento medio del residuo de las ventas marcadas); para una columna numérica, es una contribución por unidad. Cuando varias amenidades correlacionadas se estiman *conjuntamente* (p. ej., la superficie de las construcciones auxiliares junto con un indicador de taller), la regresión reparte el crédito entre ellas en lugar de contar dos veces los mismos dólares en dos filas de la descomposición. Las estimaciones respaldadas por menos de unas diez observaciones son solo indicativas y deben corroborarse mediante el costo depreciado o una muestra ampliada (p. ej., indicadores recuperados de los comentarios públicos); la propia restricción residual acota cualquier valor que finalmente se asigne. earthUI (v0.11+) implementa este método directamente como la pestaña *Residual Contribution*, incluidas la estimación conjunta, la evidencia de residuos por venta y una nota de documentación para el expediente de trabajo.

Notación: Utilizamos la notación de la tabla anterior para llegar a:

$$\epsilon_i = \epsilon_{i,1} + \epsilon_{i,2} + \dots + \epsilon_{i,n_c} \quad (11)$$

donde n_c es el número de componentes de la descomposición del residuo.

El VE debe emplear su juicio para descomponer el residuo en los componentes dados. Podría pensarse que esto permitiría que el sesgo se introdujera en la conclusión de valor final. Sin embargo, no es así, porque contamos con una demostración matemática de que, si los ajustes de la descomposición suman el residuo, no afectarán la conclusión de valor. El propósito de la descomposición no es establecer el valor, sino explicar por qué el residuo es el que es en comparación con los residuos de venta de otros inmuebles. El único punto donde el sesgo del ingeniero puede afectar la conclusión de valor final es la estimación del residuo del inmueble a valuar, dado que, por supuesto, no existe un precio de venta real del inmueble a valuar con el cual calcular su residuo.

Cada inmueble tiene su conjunto de contribuciones al valor residuales, pero estas deben fusionarse en un único conjunto C sobre todos los inmuebles:

Definiendo: $A_i = \{\{\epsilon_{i,c}\}_{c=1}^{n_c}\}$

$$C = \{x \mid x \in \bigcup_{i=0}^{n_p} A_i\}$$

donde $[\epsilon_{i,c}]$ es el nombre de variable para el valor $\epsilon_{i,c}$, A_i es el conjunto de dichas variables y n_c es el número de valores en cada ϵ_i .

Así, con los datos del inmueble a valuar y de los comparables en columnas separadas, debajo de todas las contribuciones al valor de las variables del modelo de regresión para cada inmueble, deberíamos ver el conjunto C de n_c contribuciones al valor residuales correspondientes.

9. Demostraciones fundamentales del RCA

Sea:

$$\begin{aligned} n_p &= \# \text{ of properties} \\ n_r &= \# \text{ of regression variables} \\ n_c &= \# \text{ of residual component variables} \end{aligned}$$

Demostración I: Todos y cada uno de los precios de venta ajustados de los comparables son iguales al precio de venta estimado del inmueble a valuar.

Para cualquier comparable i , el precio de venta ajustado α_i es la suma del precio de venta neto π_i más los ajustes: las diferencias entre las contribuciones al valor del inmueble a valuar y las del comparable para el valor base, las variables de regresión $j = 1, 2, \dots, n_r$ y el residuo. Para el inmueble a valuar, que carece de una venta real, el precio de venta neto π_0 se define como la conclusión de valor $\beta + \varphi_{0,1} + \varphi_{0,2} + \dots + \varphi_{0,n_r} + \epsilon_0$. Entonces:

$$\begin{aligned} \alpha_i &= \pi_i + (\beta - \beta) + (\varphi_{0,1} - \varphi_{i,1}) \\ &\quad + (\varphi_{0,2} - \varphi_{i,2}) + \dots + \\ &\quad (\varphi_{0,n_r} - \varphi_{i,n_r}) \\ &\quad + (\epsilon_0 - \epsilon_i) \\ &= \pi_i + (\beta + \varphi_{0,1} + \varphi_{0,2} + \dots \\ &\quad + \varphi_{0,n_r} + \epsilon_0) \\ &\quad - (\beta + \varphi_{i,1} + \varphi_{i,2} + \dots + \\ &\quad \varphi_{i,n_r} + \epsilon_i) \\ &= \pi_i + \pi_0 - \pi_i \\ &= \pi_0 \end{aligned} \tag{12}$$

Demostración II: Alterar la descomposición del residuo, bajo la restricción de que todas las contribuciones al valor de los componentes residuales sumen el residuo del comparable, no tiene ningún impacto en la conclusión de valor final del protocolo RCA.

Esto es casi trivial. Sin embargo, si no se reflexiona sobre la matemática, resulta fácil pasarlo por alto, porque los valuadores tradicionales se concentran en los ajustes en lugar de en las contribuciones al valor. Son las contribuciones al valor las que se cancelan aquí — una omisión de décadas. Para algún comparable k , $k > 0$, el ajuste del residuo total es igual a la suma de los n_c componentes residuales de la descomposición, que a su vez es igual a la suma de las diferencias de las contribuciones al valor, la cual es igual al residuo total del inmueble a valuar menos el residuo total del comparable. Y esto es independiente del número de componentes residuales

creados o de lo que representen. Mientras el valor de los componentes residuales sea igual al residuo total tanto para el inmueble a valorar como para el comparable elegido, ξ_k no cambiará.

$$\begin{aligned}
 \xi_k &= \xi_{k,1} + \xi_{k,2} + \dots + \xi_{k,n_c} \\
 &= (\epsilon_{0,1} - \epsilon_{k,1}) + (\epsilon_{0,2} - \epsilon_{k,2}) + \dots + \\
 &\quad (\epsilon_{0,n_c} - \epsilon_{k,n_c}) \\
 &= (\epsilon_{0,1} + \epsilon_{0,2} + \dots + \epsilon_{0,n_c}) - \\
 &\quad (\epsilon_{k,1} + \epsilon_{k,2} + \dots + \epsilon_{k,n_c}) \\
 &= \epsilon_0 - \epsilon_k
 \end{aligned} \tag{13}$$

10. Modelo MARS

10.1. Funciones “ g ” de Variables del Modelo y de Interacción

Un principio fundamental del método RCA es que el “razonamiento explicativo” reviste la máxima importancia. Lamentablemente, proporcionar gráficas significativas para las diversas ecuaciones de contribución al valor, en especial cuando se emplean interacciones y variables factor, puede resultar difícil. Por ello se considera imperativo establecer un protocolo claro y estricto para la creación de estas gráficas. Eso se hace aquí, pero debe entenderse que el protocolo está dirigido al desarrollador que escribe el código en R o Python para generar las gráficas, y podría ser de interés solo superficial para quienes no participan en la programación.

El modelo de precios generado por MARS es la suma de una constante base más términos adicionales, cada uno de los cuales involucra de 1 a 3 variables. Para analizarlo, primero se divide el modelo en términos individuales. Luego, los múltiples términos que comparten el mismo conjunto de variables, y que representan una sola interacción, se combinan en expresiones de interacción únicas. Supóngase que el grado máximo de regresión es 3. Considérese un modelo de regresión de tercer grado compuesto por una constante base, n_1 expresiones de primer grado, n_2 expresiones de interacción de segundo grado y n_3 expresiones de interacción de tercer grado. Esto nos da cuatro grupos distintos de expresiones, que son de una sola variable o interacciones de múltiples variables, cada uno graficado de manera diferente según dos factores: (1) si incluyen una variable factor (una variable categórica), y (2) el número de variables únicas que contienen. Específicamente:

- Las expresiones de primer grado (1 variable) se grafican en 2D.
- Las expresiones de segundo grado (2 variables) se grafican en 3D.
- Las expresiones de tercer grado (3 variables) se grafican como un *conjunto* de gráficas 3D.
- Si una expresión contiene una variable factor, cada factor único reduce en uno la dimensionalidad de la gráfica (p. ej., una gráfica 3D se convierte en 2D si una de las variables es un factor).

La indexación de cada función g proporciona la información necesaria para determinar cómo debe graficarse, reflejando el número de variables únicas y la presencia de variables factor.

10.2. Indexación de las Funciones g

Cada función g se indexa como $f_{j,k}g_k^j$, donde:

- $f_{j,k}$ es el primer índice (un superíndice antes de g):
 - j va de 0 a 3 e identifica el grupo:
 - $j = 0$: la constante base (una excepción a la indexación habitual).
 - $j = 1$: expresiones de primer grado.
 - $j = 2$: expresiones de segundo grado.
 - $j = 3$: expresiones de tercer grado.
 - k es el índice de una función específica dentro de su grupo (p. ej., $k = 1$ para la primera función del grupo j).
 - El valor de $f_{j,k}$ indica el número de variables factor únicas en la función:

- $f_{j,k} = 0$: sin variables factor.
- $f_{j,k} > 0$: el número de variables factor (p. ej., 1, 2 o 3).
- Índices posteriores en g :
 - El superíndice j después de g (p. ej., g^j) repite el número de grupo para mayor claridad.
 - El subíndice k después de g (p. ej., g_k) coincide con el k de $f_{j,k}$, identificando la función dentro de su grupo.

Si bien los índices de f y g coinciden numéricamente, sus funciones difieren: $f_{j,k}$ cuenta las variables factor, lo cual es fundamental para las decisiones de graficación, mientras que j y k en g definen su grupo y su posición.

10.3. Dimensiones de Graficación

La dimensión d de la gráfica para una función $f_{j,k} g_k^j$ se calcula como:

$$d = j - f_{j,k} \quad (14)$$

donde:

- j es el número de variables únicas (o número de grupo, sin considerar el caso de la constante).
- $f_{j,k}$ es el número de variables factor.
- d es el número de variables continuas (no factor) por graficar, con un máximo de 3 (dado que el grado máximo es 3). Una función con d variables continuas se dibuja en $d + 1$ dimensiones: $d = 1$ da una gráfica 2D, $d = 2$ una gráfica 3D (o un mapa de calor 2D), y $d = 3$ un *conjunto* de gráficas 3D.

Por ejemplo:

- Si $j = 3$ (3 variables) y $f_{j,k} = 0$ (sin variables factor), entonces $d = 3 - 0 = 3$, lo que requiere un conjunto de gráficas 3D, cada una fijando una de las variables.
- Si $j = 3$ y $f_{j,k} = 1$ (una variable factor), entonces $d = 3 - 1 = 2$, lo que requiere una gráfica 3D (o un mapa de calor 2D) para cada valor de la variable factor que aparece en el modelo.
- Si $j = 0$ (la constante), normalmente no se necesita graficación, pues se trata de un solo valor.

Cuando $d = 3$ (p. ej., una expresión de tercer grado sin variables factor), visualizar la función completa puede requerir múltiples gráficas 3D, cada una fijando una variable en distintos valores para mostrar un subconjunto del comportamiento de la función.

10.4. La Ecuación Final del Modelo

El modelo MARS se descompone en cuatro grupos de expresiones (constante, de primer, segundo y tercer grado). La indexación $f_{j,k} g_k^j$ nos indica:

- j : el grupo y el número de variables.
- $f_{j,k}$: el número de variables factor.

- $d = j - f_{j,k}$: la dimensión de la gráfica.

Esta estructura guía la forma en que se visualiza cada función g , garantizando que el método de graficación corresponda a la complejidad de la expresión y a los tipos de variables.

La estimación del precio de venta $\hat{P}$ proporcionada por el modelo MARS puede expresarse como:

$$\begin{aligned} \hat{P} = & f_{0,1}g_1^0 + \sum_{q=1}^{n_1} f_{1,q}g_q^1(x_*) \\ & + \sum_{r=1}^{n_2} f_{2,r}g_r^2(x_*, y_*) \\ & + \sum_{s=1}^{n_3} f_{3,s}g_s^3(x_*, y_*, z_*) \end{aligned} \quad (15)$$

$n_1 = \#$ of first degree terms
 $n_2 = \#$ of second degree terms
 $n_3 = \#$ of third degree terms

10.5. Número de Gráficas Necesarias para Graficar un Modelo

¿Cuántas gráficas se requieren para visualizar todas las funciones de un modelo? La respuesta depende de las variables, de sus interacciones y de si son variables factor o no factor.

El VE puede limitar las variables y las interacciones en el archivo de configuración de Excel para R/earth, lo que influye en qué interacciones se permiten. Un enfoque práctico consiste en dejar que R/earth evalúe todas las interacciones posibles y seleccione aquellas con contribuciones significativas a la estimación del precio de venta, verificando al mismo tiempo problemas como el sobreajuste o la colinealidad. Algunas interacciones pueden quedar excluidas, lo cual afecta a las interacciones de grado superior (p. ej., si se excluye un par en una interacción de tercer grado, se excluye la combinación completa).

Para las variables factor (p. ej., MLS Area con 12 valores interactuando con GLA), la graficación se vuelve más compleja. Esto podría requerir 12 gráficas—una por zona. Si otro factor como Architectural Style (6 valores) interactúa con MLS Area y GLA, podrían necesitarse hasta $12 \cdot 6 = 60$ gráficas. Sin embargo, R/earth rara vez encuentra significativas todas las combinaciones; por lo general, solo unas pocas interacciones clave (p. ej., zonas y estilos específicos) ameritan graficarse, según lo identifica el propio modelo.

He aquí una guía simplificada del número de gráficas necesarias:

1. **Funciones de una sola variable (no factor):** 1 gráfica.
2. **Funciones de una sola variable (factor):** 1 gráfica de barras que combina todos los valores del factor.
3. **Función base:** No se necesita gráfica.
4. **Interacciones de 2.º grado (sin factores):** 1 gráfica (p. ej., mapa de calor 2D con colores para el precio de venta).
5. **Interacciones de 2.º grado (1 factor):** 1 gráfica por cada valor del factor en el modelo (p. ej., 2 valores = 2 gráficas).
6. **Interacciones de 2.º grado (2 factores):** 1 gráfica con una matriz de combinaciones de factores (p. ej., 3 zonas $\times$ 2 estilos = 6 celdas).
7. **Interacciones de 3.º grado (sin factores):** Múltiples gráficas (p. ej., 4 gráficas para el número de baños = 1, 2, 3, 4).

8. **Interacciones de 3.º grado (1 factor):** 1 gráfica por cada valor del factor en el modelo.
9. **Interacciones de 3.º grado (2 factores):** 1 gráfica por cada combinación de valores de los factores en el modelo.
10. **Interacciones de 3.º grado (3 factores):** 1 gráfica por cada valor del factor con el menor número de valores distintos, usando una matriz de colores para los otros dos.

Escenarios más complejos pueden requerir técnicas de graficación avanzadas.

Existen otras posibilidades con gráficas probablemente más sofisticadas.

Un ejemplo de un modelo completo generado por MARS se encuentra en [Figure 1](#).

11. Características de los datos

La siguiente sección analiza MARS. MARS depende por completo de los datos que recibe, y la mayor parte de estos datos provendrá de algún MLS (Multiple Listing Service) o de los Tax Assessors. Sin embargo, pueden pasar por otros proveedores de datos que los mantienen y mejoran para luego venderlos con cierta ganancia. En algunas zonas son bastante confiables; en otras, no tanto. En casos críticos, el inspector del inmueble debe realizar sus propias observaciones y mediciones, cuya exactitud puede cuestionarse.

El Valuation Engineer necesita comprender qué representan los datos y qué tan bien cumplen la función de representar lo que se supone que representan. Esto plantea toda una lista de cuestiones:

- ¿Cómo se clasifica una variable como una medición?
- Si una variable no es una medición, ¿qué otra cosa puede ser?
- ¿Cómo puede ser útil la regresión si los datos que recibe no son exactos?

La mayor parte de los datos que obtenemos de los servicios de listings de la MLS y de los datos del Tax Assessor para la regresión son datos numéricos que implican mediciones o conteos. Una parte es lógica, verdadero o falso.

11.1. Propósito del residuo

El residuo sirve como un indicador indirecto del valor aportado por las variables no incluidas en el modelo de regresión. En el lenguaje de la literatura hedónica, estas características no medidas son *variables latentes*: variables que influyen en el precio de venta pero que no se miden directamente, en la mayoría de los casos porque no existe una forma práctica de medirlas objetivamente. Sus valores deben, en cambio, inferirse y, en la práctica, el valor de una variable latente se estima a menudo mediante juicio subjetivo — en RCA, un juicio anclado en el ranking de residuos. Al ordenar los inmuebles según su residuo o su residuo por pie cuadrado, obtenemos un ranking objetivo que va desde los inmuebles con menor atractivo hasta los de mayor atractivo. Este método de clasificación es eficaz, siempre que un analista competente con un conocimiento profundo del mercado local construya el modelo MARS (Multivariate Adaptive Regression Splines). Los siguientes son algunos requisitos adicionales:

- Calidad de los datos y del modelo: Un modelo de regresión bien construido y datos de alta calidad son requisitos previos para obtener residuos confiables. En la San Francisco Bay Area, mi objetivo es desarrollar un modelo con un R^2 de aproximadamente 75-80 % y un R^2 de validación cruzada (CV R^2) de 55-70 %.
- Evitar el sobreajuste, limitar el R^2 : El Valuation Engineer debe aspirar a un R^2 máximo de alrededor del 80 % en la mayoría de los mercados maduros de zonas de precios más altos, como la SF Bay Area, porque gran parte del valor competitivo de un inmueble residencial reside en su atractivo subjetivo, que no es cuantificable y, por lo tanto, no puede incluirse directamente en los modelos de regresión. Un R^2 significativamente superior al 80 % podría indicar un modelo sobreajustado, lo que exige un examen más detenido. No obstante, una advertencia: existen zonas de mercado con valores de R^2 superiores al 90 % sin sobreajuste, y hay zonas difíciles de valorar donde lo mejor que puede lograrse es un 50-60 % o menos.
- R^2 alto en zonas específicas: Sin embargo, hay regiones, como fraccionamientos u otras zonas residenciales uniformes, donde valores más altos de R^2 son sorprendentemente comunes debido a la homogeneidad de los inmuebles.

11.2. Las calificaciones de estado y calidad como regresores aproximados

Las escalas de calificación de estado y calidad de Fannie Mae (C1–C6 y Q1–Q6) merecen un comentario especial. Estas escalas son demasiado burdas para tratarse como verdaderas características medidas y, dado que las asigna el valuador, son en la práctica calificaciones subjetivas — variables latentes por derecho propio. Resulta útil distinguir explícitamente dos niveles: la calificación burda del valuador utilizada como regresor, *estado de grano grueso* (condition-1) y *calidad de grano grueso* (quality-1), frente a las variables latentes de grano fino recuperadas del residuo, *estado de grano fino* (condition-2) y *calidad de grano fino* (quality-2). El VE puede aceptar las calificaciones de grano grueso del valuador como aproximaciones y efectuar la regresión sobre ellas, obteniendo contribuciones al valor aproximadas que resultan útiles como asignaciones de valor iniciales para las variables latentes de grano fino. Cuando así se procede, el protocolo consiste en restar de la estimación del modelo las contribuciones al valor estimadas por la regresión para estas dos variables de grano grueso — retirándolas de los atributos medidos en la sección de regresión del Sales Grid — y sumarlas de nuevo al residuo. A partir de ahí, se retiran posteriormente del residuo y se asignan a las variables latentes de grano fino correspondientes, `res_condition1` y `res_quality1`. La información adicional extraída de los comentarios públicos de la MLS (`public_remarks`) y de listas de cadenas relacionadas se utiliza luego para realizar microrrefinamientos de estos valores de las variables latentes. El resultado final es que las contribuciones al valor de grano fino reemplazan a las contribuciones al valor de grano grueso: las calificaciones burdas sirven únicamente como andamiaje para el modelo aproximado, y los valores que finalmente se reportan para estado y calidad son los de grano fino, derivados del residuo.

11.3. Errores de datos

Según comentarios en redes sociales, los datos de la Multiple Listing Service (MLS) de California son, en general, de mayor calidad que los de muchas otras regiones de Estados Unidos. Típicamente, cabe esperar que los datos de la MLS sean de buena calidad en la mayoría de las áreas metropolitanas maduras, pero de menor calidad en las zonas rurales.

Si bien ocasionalmente hay errores en los datos por diversas razones, estos errores en general deberían cancelarse entre sí y tener un impacto mínimo en los promedios utilizados en los modelos de precios si existen datos suficientes.

Sin embargo, pueden producirse sesgos sistemáticos en zonas específicas. Por ejemplo, considérese un vecindario donde muchas viviendas se construyeron inicialmente con un segundo piso completo, pero a algunos compradores se les dio la opción de eliminar habitaciones del segundo piso para crear techos abovedados en una o más habitaciones de la planta baja, lo que reduce significativamente la superficie habitable bruta (GLA). Por una u otra razón, los cambios no se actualizaron posteriormente en los registros del Tax Assessor. En consecuencia, la MLS podría mostrar 2,400 pies cuadrados cuando el valor medido es de solo 2,000 pies cuadrados. En tales escenarios, los datos de la MLS deben corregirse antes de ejecutar la regresión MARS (Multivariate Adaptive Regression Splines). Es importante señalar que obtener esta información para los comparables puede requerir cierto esfuerzo de investigación. Las discrepancias deben anotarse para referencia futura.

Dado lo anterior, es razonable suponer que, en zonas con MLS de alta calidad, los errores que se filtran al residuo son mínimos en comparación con el impacto de características subjetivas y no medidas como el diseño, el estado y la calidad. Esta es una consideración importante al estimar la exactitud del método RCA. Por ejemplo, si el valor de R^2 para una zona de mercado dada es del 80 %, entonces aproximadamente el 20 % del precio se atribuye al residuo CQA. Un error de $\pm 10\%$ al ubicar el inmueble a valuar dentro del ranking de residuos se propaga, por lo tanto,

como un error de $\pm 10\% \cdot 20\% = \pm 2\%$ en la conclusión de valor final. ¡Así es como se mejora la exactitud! En muchos casos son posibles estimaciones de valor exactas muy por debajo de $\pm 1\%$.

Por lo tanto, tratar los residuos como indicadores indirectos del valor colectivo de las variables no modeladas en la regresión MARS es razonable, siempre que el Valuation Engineer cuente con suficiente experiencia en el mercado local para identificar dónde y por qué existen sesgos sistémicos de medición. Este enfoque asegura que el análisis refleje con exactitud los valores reales de los inmuebles, reconociendo las limitaciones y los matices de los datos inmobiliarios.

Ejemplo desarrollado: medir determinantes del valor no medidos en Sea Ranch

Cuando el R^2 alcanzable en una zona de mercado es más bajo de lo que la aparente homogeneidad del mercado haría esperar, la respuesta del marco consiste en identificar y medir la variable faltante en lugar de aceptar la limitación. La comunidad de Sea Ranch, en el condado de Sonoma, lo ilustra directamente.

Un caso concreto de la respuesta de “medir, no aceptar”. La comunidad de Sea Ranch, en el condado de Sonoma, es un desarrollo costero donde la distancia al océano es un determinante del valor primario: una vivienda a varios cientos de pies del acantilado alcanza un precio sustancialmente distinto del de una situada a un cuarto de milla tierra adentro, manteniendo constantes las demás características. Los datos de la MLS para la comunidad no registran la distancia al océano como variable. Un modelo MARS inicial ajustado con las variables disponibles de la MLS produjo un poder explicativo menor del que la aparente homogeneidad del mercado debería haber sustentado, con una estructura de residuos que no ordenaba suficientemente los inmuebles según ninguna característica visible en los datos — una señal de que faltaba un determinante importante en el modelo, y no de que el mercado fuera genuinamente ruidoso.

El remedio fue la medición directa. El autor obtuvo un mapa de la comunidad que mostraba la disposición de los inmuebles respecto de la línea de costa y midió la distancia al océano de varios cientos de viviendas. Las distancias medidas se agregaron como variable al conjunto de datos de la MLS y se volvió a ejecutar la regresión MARS. La nueva variable fue seleccionada por el modelo como significativa, el poder explicativo aumentó sustancialmente y el ranking de residuos pasó a ordenar los inmuebles según patrones consistentes con las características subjetivas restantes — vista, estado, calidad — en lugar de la señal latente de distancia al océano que previamente lo había contaminado. La elevación se ha abordado de manera similar en vecindarios de ladera donde la variable de la MLS falta o no es confiable.

El principio metodológico es general. Cuando el R^2 alcanzable cae por debajo de lo que la aparente homogeneidad del mercado debería sustentar, la primera hipótesis es que falta una variable importante en el modelo, y la respuesta apropiada es identificar cuál es y recolectar las mediciones directamente. Las fuentes incluyen mapas comunitarios, datos GIS, registros del Tax Assessor, registros del departamento de construcción, imágenes satelitales e inspección de campo. El costo de este trabajo de medición es real pero acotado; una vez recolectada la variable para una zona de mercado dada, puede trasladarse a encargos futuros en esa zona. RCA absorbe nuevas variables limpiamente porque el marco es indiferente a qué variables entran en la regresión MARS — la estructura de residuos indica al VE si la nueva variable captura señal real, y las contribuciones al valor se ajustan en consecuencia.

La respuesta alternativa — aceptar un R^2 bajo y reportar bandas de error más amplias — es defendible cuando los determinantes del valor faltantes son genuinamente inmediables (preferencias específicas del comprador, sentimiento de mercado transitorio, circunstancias idiosincrásicas de la transacción), pero no cuando simplemente no se han recolectado. El primer paso del VE ante un modelo débil debe ser preguntarse si la señal faltante corresponde a lo segundo.

11.4. ¿Qué es la exactitud?

La exactitud perfecta de una valuación requiere:

1. Conformidad con el valor de mercado: La transacción de venta hipotética del inmueble a valorar se ajusta al 100 % a la definición dada de valor de mercado
2. Desarrollo de un modelo exacto y robusto: Se desarrolla un modelo MARS de alta calidad con un R^2 de alrededor de 75-80 % y un CVR2 de aproximadamente 55-70 %, según la ubicación.
3. Buen análisis de errores de datos: Todos los errores de datos probables de las variables medidas se comprenden y están contenidos, con poca filtración hacia los residuos.
4. Contabilización del sesgo sistémico: Cualquier sesgo sistémico en los datos debe contabilizarse y ajustarse.
5. Análisis del ranking de residuos: Asegurarse de que los inmuebles del ranking de residuos sigan un patrón razonable de características desde el rango más bajo hasta el más alto. Las anomalías, como ventas testamentarias, ventas cortas (short sales) y ventas en subasta, deben investigarse, explicarse y, potencialmente, eliminarse.
6. Ubicación del inmueble a valorar: Con una comprensión profunda del ranking de residuos, el inmueble a valorar debe posicionarse con exactitud. Debe identificarse un inmueble con un atractivo significativamente mayor y otro con un atractivo significativamente menor, y el inmueble a valorar debe ubicarse entre ellos y puntuarse. Luego se intenta estrechar el ranking entre el inmueble superior y el inferior tanto como sea posible.
7. Razonamiento explicativo: Nunca debe sacrificarse el razonamiento explicativo en aras de la exactitud. El valuador debe poder explicar las diferencias de valor entre el inmueble a valorar y los inmuebles comparables, al menos a una persona de inteligencia razonable. Por ejemplo, un buen modelo MARS con un R^2 y un CVR2 óptimos puede utilizarse para explicar las contribuciones al valor de las características medidas, mientras que la descomposición del residuo por parte del Valuation Engineer puede utilizarse para explicar las contribuciones al valor de otras características.
8. Sobreajuste: Deben evitarse los modelos sobreajustados, no utilizando características del inmueble que puedan identificar, ya sea individualmente o en combinación, uno o varios inmuebles. En particular, debe evitarse el uso de variables factor cuyos valores puedan especificar un número reducido de inmuebles. Una variable factor que identifica vecindarios en conjuntos de datos de tamaño considerable es apropiada, pero una que identifica inmuebles con piscina en una zona donde existen muy pocas piscinas debe usarse con cautela. Una característica que identifica inmuebles únicos, como coordenadas GIS exactas, números de parcela o direcciones, está absolutamente prohibida.

Si el R^2 de la regresión es del 80 %, esperamos que alrededor del 20 % del valor del inmueble a valorar se deba al residuo. Ahora bien, esto es solo aproximado. Sabemos por la curva CQA-Residuo que la mayor parte de la distribución de los residuos se encuentra en los extremos inferior y superior del CQA. Con base en la curva CQA-Residuo y la pendiente extrema para los valores bajos y altos de CQA, podríamos esperar una estimación menos exacta del residuo en estas zonas. Sin embargo, esto se ve contrarrestado por la relativa escasez de inmuebles en estas zonas y por la mayor diferencia de atractivo que observamos entre inmuebles adyacentes. Así, mientras que en el rango medio podríamos hablar del impacto de un rango de $\pm 5\%$, en los extremos nos preocuparía más un rango algo más estrecho de $\pm 2 - \pm 3\%$

Calcular la exactitud del RCA es, en general, algo difícil debido a la no linealidad de la curva CQA-Residuo. Pero es posible caso por caso, siempre que se disponga de un ranking de residuos razonable. El rango de las puntuaciones CQA es de 0.0 a 10.0. Con un buen modelo MARS, el ranking de residuos debería mostrar un patrón claro de menor a mayor atractivo, y no debería resultar difícil calificar dentro de un $\pm 5\%$ del rango total, es decir, ± 0.5 para la puntuación CQA. Si hay 100 inmuebles en la regresión, eso equivaldría a clasificar el inmueble a valorar con exactitud dentro de un grupo de 10 inmuebles. Así, con referencia a los datos de ventas de Burlingame de la Tabla 3, calculemos la exactitud (o el error) aproximada para un CQA estimado por el VE de 7.5, 5.1 y 1.9, donde su conclusión de valor para el inmueble a valorar es de \$3,000,000. Nótese que, dado que la Tabla 3 solo contiene una muestra de los varios cientos de inmuebles utilizados, será necesario interpolar el residuo cuando corresponda (suponiendo linealidad). Para ser claros, suponemos que el VE está muy seguro, en cada caso, de que su clasificación del inmueble a valorar se encuentra dentro de un rango de ± 5 inmuebles en el ranking de residuos de solo 100 inmuebles.

CQA= 7.5 ± 0.5					
CQA	Residuo	Dif. abs.	Error esp.	Precio de venta hip.	Error pct.
8.00	\$257,112				
7.50	\$197,113	\$59,999			
7.00	\$154,615	\$42,498	$\pm \$51,249$	\$3,000,000	$\pm 1.71\%$

CQA=5.1 ± 0.5					
CQA	Residuo	Dif. abs.	Error esp.	Precio de venta hip.	Error pct.
5.60	\$50,911				
5.10	\$19,559	\$31,352			
4.60	-\$28,207	\$47,766	$\pm \$39,559$	\$3,000,000	$\pm 1.32\%$

CQA= 1.9 ± 0.5					
CQA	Residuo	Dif. abs.	Error esp.	Precio de venta hip.	Error pct.
2.40	-\$228,411				
1.90	-\$276,381	\$47,970			
1.40	-\$326,938	\$50,557	$\pm \$49,264$	\$3,000,000	$\pm 1.64\%$

Finalmente, la única manera objetiva de verificar la exactitud es establecer un protocolo estricto (que este artículo solo insinúa parcialmente a grandes rasgos) y verificar su cumplimiento. Esto es necesario porque los precios de venta reales pueden desviarse del valor de mercado debido a imperfecciones del mercado. Comparar avalúos con ventas posteriores puede proporcionar cierta indicación de la exactitud. Sin embargo, las complejidades y la inestabilidad del mercado inmobiliario exigen la pericia de valuadores o de Valuation Engineers altamente calificados para proporcionar estimaciones de valor basadas en estándares y protocolos establecidos. Estos deben ser lo suficientemente rigurosos como para asegurar que Valuation Engineers competentes lleguen de manera independiente a casi la misma conclusión de valor. Este debería ser un objetivo alcanzable en la mayoría de los casos en que se dispone de buenos datos.

11.5. El inmueble a valorar como anomalía

Dado que estimar el valor de mercado de un inmueble requiere el análisis de transacciones de venta recientes, es justo preguntarse si las transacciones de venta recientes representan adecuadamente las diversas características de un inmueble a valorar más antiguo que carece de

modernización y reparaciones. Esto es particularmente cierto en la San Francisco Bay Area, donde las transacciones de venta recientes suelen corresponder a viviendas que han pasado por actualizaciones para mejorar su precio de venta. Así, el estado y la calidad promedio de los comparables de venta son generalmente superiores a los de viviendas más antiguas que no han estado en el mercado durante bastante tiempo. Esta diferencia de valor debería ser captada por la estimación del residuo para el inmueble a valorar o, en otras palabras, por su puntuación CQA. Esta es una consideración importante con respecto a la calidad de los datos y la exactitud al valorar el inmueble a valorar.

11.6. La importancia del CVR2 frente a datos inesperados

La regresión MARS se basa en múltiples iteraciones de R/earth aplicadas a un subconjunto de inmuebles seleccionado aleatoriamente, que típicamente comprende del 80 % al 90 % del conjunto de datos de la MLS, denominado conjunto de entrenamiento. El 10 % - 20 % restante de los inmuebles sirve como conjunto de prueba. En esta fase, el modelo MARS se crea a partir del conjunto de datos de entrenamiento y luego se utiliza para predecir los precios de venta de los inmuebles del conjunto de datos de prueba, que después se comparan con los precios de venta reales. Esta metodología ha sido reconocida por los especialistas en minería de datos como la estrategia más eficaz para desarrollar un modelo robusto capaz de predecir los precios de venta de inmuebles aún no observados.

Sin embargo, la utilidad primaria de la métrica CVR2 no radica en reflejar la exactitud fáctica del modelo de regresión respecto de las ventas reales de la zona de mercado, sino en optimizar las capacidades predictivas del modelo respecto de ventas hipotéticas de inmuebles a una fecha efectiva especificada. Estos inmuebles no están incluidos en el conjunto de datos de entrenamiento y pueden exhibir combinaciones de características nuevas o inesperadas, ausentes tanto en los datos de los inmuebles del conjunto de entrenamiento como en los del conjunto de prueba.

Dado este contexto, ¿cómo podemos estimar el precio de venta de un inmueble que no ha sido actualizado en los últimos 15 años, suponiendo que fue puesto en venta, por ejemplo, un mes antes de la fecha efectiva del avalúo? Un modelo que demuestra un CVR2 sólido tiene más probabilidades de proporcionar la estimación más exacta que un modelo que no ha pasado por esta validación cruzada repetitiva. También es vital que el conjunto de datos incluya un número suficiente de inmuebles comparables con distintos grados de actualizaciones, reparaciones y modernización.

En última instancia, si poseemos un modelo robusto que predice con exactitud las contribuciones al valor de las diversas características de los inmuebles en nuestro conjunto de datos de inmuebles comparables, podemos ubicar eficazmente el nuevo inmueble a valorar dentro del ranking por residuo que el modelo establece para estos inmuebles. Esta ubicación nos permite asignar al inmueble una puntuación CQA y un residuo apropiados.

12. Residuos

Como lo indica el nombre “Residual Constraint Approach,” los residuos son centrales en esta metodología. Aquí, residuo se refiere a la diferencia entre el precio de venta neto de un inmueble y el precio de venta predicho por el análisis de regresión.

Un residuo positivo sugiere que el inmueble se vendió por más que su precio estimado, con base en atributos medibles típicos como la superficie habitable bruta (GLA) y el tamaño del lote, lo que implica que podría tener un atractivo mayor que el promedio. Un residuo negativo indica que el inmueble se vendió por menos, lo que potencialmente sugiere un menor atractivo u otros atributos negativos no capturados.

Dado que las viviendas más grandes probablemente tienen residuos más grandes, podríamos utilizar el residuo por pie cuadrado con base en la GLA para una comparación más normalizada. Una discusión más profunda de este tema se reserva para otro artículo.

12.1. Ranking & Scoring (clasificación y puntuación)

Después de crear el modelo MARS (Multivariate Adaptive Regression Splines) para las variables medidas, se calculan los residuos para cada inmueble. Estos residuos representan la diferencia entre los precios de venta netos observados y los precios de venta predichos por el modelo. Debe crearse una columna adicional para el Residual/SF, es decir, el residuo dividido por la superficie habitable de la residencia, en el entendido de que, en general, el residuo de inmuebles específicos es más significativo con el tamaño de la superficie habitable. Otras posibles estadísticas pueden funcionar mejor en algunos vecindarios, pero son tema para otra ocasión.

Supuesto 7: “Clasificación por residuo”

En este artículo, las referencias a “clasificación por residuo” deben entenderse como clasificación por residuo, por residuo/SF, o clasificación por alguna otra función de los residuos, según el Valuation Engineer lo considere más apropiado para la zona de mercado. La experiencia del autor en el SF Bay Area indica que el residuo/SF es generalmente el más confiable, especialmente cuando se trata de un inmueble a valorar con una GLA (Gross Living Area) inusualmente pequeña o grande.

A continuación, los inmuebles se ordenan por sus residuos, del más negativo al más positivo. Luego se genera una puntuación de estado, calidad y atractivo (Condition-Quality-Appeal, CQA) para cada inmueble, en un rango de 0.0 a 10.0. Esta puntuación se calcula de la siguiente manera:

La puntuación CQA corresponde al rango percentil del residuo de un inmueble entre todos los inmuebles. Específicamente, la puntuación se calcula determinando el porcentaje de inmuebles que tienen residuos menores que el del inmueble dado y dividiendo ese porcentaje entre 10. Por ejemplo, si el residuo de un inmueble es mayor que los residuos del 50 % de los inmuebles, entonces su puntuación sería 5.0.

Debido al redondeo, un inmueble puede recibir una puntuación de 10.0, que representa al inmueble con el residuo más alto. Sin embargo, una puntuación perfecta de 10.0 es teóricamente imposible porque, lógicamente, ningún inmueble puede tener un residuo mayor que el del 100 % de los inmuebles, incluido él mismo. Por otro lado, la puntuación más baja posible de 0.0 sí es posible, ya que al menos un inmueble no tiene otros inmuebles con residuos menores. Nótese que más de un inmueble podría tener una puntuación de 0.0 debido al redondeo si se comparan muchos inmuebles.

Por lo tanto, la puntuación CQA proporciona una medida intuitiva de qué tan bien los atributos

de un inmueble se alinean con las expectativas del modelo, donde las puntuaciones más altas indican una mejor alineación con las predicciones del modelo.

Es esencial reconocer que la puntuación CQA solo puede considerarse un indicador indirecto del atractivo de un inmueble comparable cuando las condiciones de venta son acordes con su valor de mercado. El Valuation Engineer es responsable de determinar si el valor residual incorpora elementos por encima o por debajo del mercado, mediante una explicación exhaustiva y una investigación minuciosa de la transacción de venta. Si bien las ventas significativamente por encima o por debajo del mercado no son frecuentes, pueden ocurrir en circunstancias tales como ventas forzadas, sucesiones testamentarias, incompetencia, colusión, etc.

La Tabla 3 presenta un extracto semirrealista de datos de la MLS, junto con los residuos y las puntuaciones CQA, tanto en total como por pie cuadrado. El número total de inmuebles analizados en este estudio fue de aproximadamente 600. Es digno de mención que varias variables que fueron incluidas en la regresión no se muestran. En otras palabras, los residuos presentados son posteriores a que el impacto de otras variables, como la fecha de venta, haya sido considerado por la regresión MARS.

Cuadro 3: Comparables de venta ordenados por residuo

SalePrice	Estimated SP	Residual	CQA	Residual/SF	CQA SF	GLA	SaleDate
\$3,450,000	\$2,318,865	\$1,131,135	10	\$420.50	9.98	2,690	2021-04
\$4,360,757	\$3,399,922	\$960,835	9.9	\$290.28	9.82	3,310	2021-09
\$4,000,000	\$3,134,444	\$865,556	9.8	\$234.57	9.60	3,690	2020-11
\$5,850,000	\$5,150,966	\$699,034	9.7	\$109.91	8.35	6,360	2021-12
\$3,600,000	\$3,023,366	\$576,634	9.5	\$269.46	9.79	2,140	2022-06
\$4,425,000	\$3,987,873	\$437,127	9.2	\$127.44	8.68	3,430	2021-08
\$2,530,000	\$2,164,564	\$365,436	8.9	\$180.02	9.28	2,030	2019-12
\$4,250,000	\$3,884,719	\$365,281	8.8	\$87.18	7.91	4,190	2021-08
\$3,350,000	\$3,034,613	\$315,387	8.5	\$130.87	8.75	2,410	2021-09
\$3,800,000	\$3,522,028	\$277,972	8.2	\$77.43	7.72	3,590	2020-09
\$4,200,000	\$3,963,748	\$236,252	7.8	\$56.12	6.89	4,210	2021-01
\$2,900,000	\$2,702,887	\$197,113	7.5	\$89.19	7.92	2,210	2021-07
\$4,250,000	\$4,103,885	\$146,115	6.9	\$39.38	6.29	3,710	2022-08
\$5,000,000	\$4,864,958	\$135,042	6.8	\$28.25	5.85	4,780	2021-07
\$4,320,000	\$4,199,672	\$120,328	6.5	\$33.90	6.04	3,550	2022-07
\$2,600,000	\$2,479,904	\$120,096	6.5	\$60.96	7.00	1,970	2021-10
\$3,400,000	\$3,301,146	\$98,854	6.2	\$30.32	5.90	3,260	2021-06
\$2,000,000	\$1,901,987	\$98,013	6.2	\$73.14	7.54	1,340	2019-11
\$3,000,000	\$2,919,718	\$80,282	5.9	\$26.50	5.79	3,030	2020-05
\$6,750,000	\$6,699,089	\$50,911	5.6	\$7.75	5.20	6,570	2020-08
\$5,652,000	\$5,601,570	\$50,430	5.5	\$12.83	5.37	3,930	2022-05
\$2,600,000	\$2,579,082	\$20,918	5.2	\$10.51	5.33	1,990	2022-03
\$1,925,000	\$1,905,441	\$19,559	5.1	\$11.51	5.36	1,700	2021-01
\$3,200,000	\$3,209,500	-\$9,500	4.8	-\$3.21	4.88	2,960	2020-11
\$4,150,000	\$4,187,560	-\$37,560	4.5	-\$10.49	4.68	3,580	2021-04
\$2,700,000	\$2,776,510	-\$76,510	3.9	-\$23.61	4.14	3,240	2021-07
\$3,450,000	\$3,567,668	-\$117,668	3.4	-\$31.05	3.90	3,790	2020-09
\$1,975,000	\$2,109,255	-\$134,255	3.3	-\$55.48	3.24	2,420	2020-02
\$1,950,000	\$2,085,431	-\$135,431	3.2	-\$68.40	2.86	1,980	2021-02
\$3,695,000	\$3,859,154	-\$164,154	3.1	-\$47.04	3.56	3,490	2021-05
\$2,288,000	\$2,455,747	-\$167,747	2.9	-\$56.86	3.19	2,950	2020-06
\$2,635,000	\$2,802,781	-\$167,781	2.9	-\$56.87	3.18	2,950	2020-09
\$2,050,000	\$2,217,794	-\$167,794	2.9	-\$82.66	2.46	2,030	2021-07
\$1,550,000	\$1,731,804	-\$181,804	2.7	-\$130.79	1.32	1,390	2021-05
\$2,800,000	\$3,028,411	-\$228,411	2.4	-\$81.58	2.51	2,800	2021-02
\$3,700,000	\$3,976,381	-\$276,381	1.9	-\$76.99	2.62	3,590	2021-08
\$2,517,500	\$2,834,974	-\$317,474	1.5	-\$120.71	1.48	2,630	2019-11
\$2,550,000	\$2,905,330	-\$355,330	1.1	-\$115.74	1.64	3,070	2019-12
\$1,815,000	\$2,263,944	-\$448,944	0.7	-\$273.75	0.15	1,640	2021-11
\$3,711,000	\$4,683,454	-\$972,454	0	-\$226.68	0.37	4,290	2021-12
\$2,225,000	\$3,332,960	-\$1,107,960	0	-\$370.56	0.03	2,990	2021-08
\$3,550,000	\$4,746,489	-\$1,196,489	0	-\$274.42	0.14	4,360	2022-06

12.2. Curva CQA-Residuo

Un gráfico de la CQA frente al Residual/SF para un vecindario invariablemente dará como resultado la curva característica CQA-Residuo que se muestra en [Figure 2](#). Este gráfico en particular representa datos de Burlingame, CA, hacia 2022. Generalmente se observan curvas similares al graficar las puntuaciones CQA frente a los valores de los residuos; sin embargo, prefiero utilizar el Residual/SF por su relevancia para contextualizar la CQA con respecto a la superficie habitable bruta (GLA) del inmueble a valorar. Al multiplicar el Residual/SF por la GLA, se puede convertir la CQA de un inmueble en un valor residual correspondiente.

El Valuation Engineer debe evaluar cuidadosamente si el tamaño del inmueble, específicamente la superficie habitable bruta (GLA), influye significativamente en la magnitud de los residuos. Esta evaluación es crítica porque puede determinar el método más exacto para calcular los valores de los inmuebles. Por ejemplo, los electrodomésticos de cocina de alta gama y los accesorios de baño lujosos pueden aumentar sustancialmente el valor residual de una vivienda. Sin embargo, en vecindarios donde el tamaño y la distribución de las cocinas y los baños son generalmente uniformes, estas mejoras podrían no correlacionarse fuertemente con la GLA. Su impacto en el valor residual puede depender menos del espacio que ocupan y más de la calidad que representan.

Por el contrario, las características y los materiales exteriores costosos, que mejoran notablemente el atractivo exterior de un inmueble, pueden tener una relación más directa con el tamaño del inmueble. Las viviendas más grandes pueden dar cabida a exteriores más elaborados, como un paisajismo extenso, materiales de techado superiores o patios amplios, que naturalmente realzan el atractivo general del inmueble y, en consecuencia, su valor residual.

Por lo tanto, es esencial que los Valuation Engineers consideren estas dinámicas al analizar los residuos. No solo deben evaluar el valor intrínseco añadido por accesorios y características de alta calidad, sino también comprender cómo estas mejoras interactúan con las dimensiones físicas del inmueble. Este enfoque integral asegura que las valuaciones reflejen con exactitud tanto los atributos cualitativos como los cuantitativos de un inmueble.

En consecuencia, se recomienda graficar tanto la curva CQA-Residuo como la curva CQA-Residual/SF y estudiar las diferencias, las cuales pueden indicar patrones en una zona de mercado. El Valuation Engineer puede hacer que su programa calcule ambos residuos y tome algún promedio ponderado para calcular el residuo que se asignará al inmueble a valorar. Puede que ello marque una gran diferencia, o puede que no.

12.3. Características de la curva CQA

La curva de estado, calidad y atractivo (CQA) ofrece un modelo sofisticado para comprender cómo fluctúan los valores de los inmuebles en función del atractivo, el estado y la dinámica del mercado dentro de un área determinada. Este modelo captura la interacción de los factores económicos y físicos que influyen en el mercado inmobiliario, reflejando las realidades matizadas de la valuación inmobiliaria en los distintos segmentos del mercado.

12.3.1. Análisis de los componentes de la curva CQA

1. Decaimiento exponencial en el lado izquierdo de la curva:

A medida que la puntuación CQA disminuye de alrededor de 2 a 0, los inmuebles exhiben características de deterioro rápido. Esta parte de la curva se representa bien mediante una función de decaimiento exponencial, lo que enfatiza la rapidez con que un inmueble puede perder su atractivo y su valor sin el mantenimiento y las actualizaciones adecuados. La

degradación física, como la madera podrida, los colores desvanecidos y el metal oxidado, acelera este declive, particularmente en viviendas que no reciben mantenimiento regular.

2. Incremento lineal en el rango medio:

La mayor parte del mercado típicamente encaja en esta categoría, donde el atractivo aumenta de manera casi lineal desde puntuaciones de alrededor de 2 hasta 8. Esta sección refleja inmuebles que en general se mantienen conforme al estándar, pero que no necesariamente presentan cualidades o ubicaciones excepcionales. Representa la amplia clase media de la vivienda, abarcando la mayoría del mercado habitacional, donde las viviendas tienen precios competitivos basados en sus características estándar y su estado general.

3. Crecimiento exponencial en el lado derecho de la curva:

Aquí, la curva se eleva exponencialmente desde las puntuaciones 8 a 10, reflejando el extremo de mayor atractivo del mercado. Los inmuebles de este segmento a menudo pertenecen a personas más adineradas que no solo invierten en un mantenimiento superior, sino también en mejoras que aumentan significativamente el atractivo y el valor del inmueble. El crecimiento exponencial refleja la escasez de inmuebles de tan alta calidad y la alta demanda entre los compradores acaudalados. Y entiéndase: una vivienda de menor precio puede calificar muy alto en términos de atractivo, mientras que una vivienda grande y costosa puede calificar bajo. Estas dinámicas se descubren mediante el uso experto de MARS y usualmente pasan desapercibidas para la mayoría de los valuadores.

12.3.2. Dinámica del mercado e implicaciones sociales

La curva CQA narra implícitamente la historia de la riqueza relativa y la disparidad económica dentro de las distintas regiones. En lugares como el condado de San Mateo, la variación en los valores de los inmuebles entre áreas como Burlingame y vecindarios más adinerados como Hillsborough, Woodside o Atherton ilustra divisiones socioeconómicas significativas. Esta disparidad afecta todo, desde el mantenimiento de los inmuebles hasta las valuaciones de mercado, y en última instancia influye en la demografía de los compradores y en las tendencias del mercado inmobiliario.

Además, la anécdota sobre una persona relativamente adinerada que invierte en una vivienda de alta calidad dentro de un vecindario de nivel medio con el que se siente cómoda y donde, por diversas razones, preferiría vivir, subraya otro aspecto crítico de la economía inmobiliaria — el concepto de inmuebles “sobreconstruidos” (overbuilt). Tales inmuebles, aunque muy atractivos, pueden no encontrar compradores locales capaces de pagar el precio que podrían haber alcanzado en una zona de mercado de precios más altos.

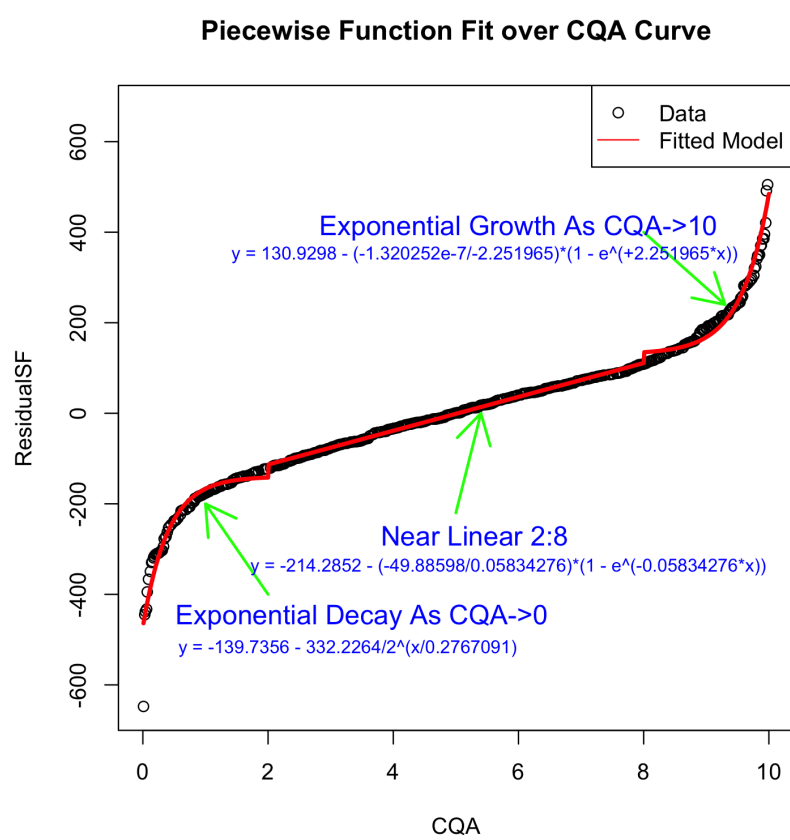
La curva CQA-Residuo es característica de cada zona de mercado, y no debe pasarse por alto al realizar el análisis.

12.4. Residuo del inmueble a valorar

A diferencia de los comparables, el inmueble a valorar no posee un residuo determinado objetivamente; este debe ser estimado por el Valuation Engineer (VE). Esta estimación introduce un grado de subjetividad en el protocolo del Residual Constraint Approach (RCA), marcando una coyuntura crítica en la que el juicio personal desempeña un papel fundamental.

Para estimar la puntuación de estado, calidad y atractivo (CQA), el VE analiza meticulosamente los inmuebles clasificados, buscando patrones discernibles dentro de las clasificaciones. Este análisis a menudo implica una revisión detallada de las fotos de la MLS asociadas con cada

Figura 2: Función característica CQA-Residuo



inmueble. Una revisión exhaustiva es impracticable para un conjunto de datos grande, como uno de 600 inmuebles. En su lugar, el VE aplica un razonamiento práctico para aproximar la posición probable del inmueble a valorar dentro de la clasificación general. Los inmuebles que requieren reparaciones extensas, o “fixers,” generalmente se posicionan cerca del fondo, mientras que aquellos con actualizaciones y atractivo superiores tienden a clasificarse cerca de la cima.

El VE adopta entonces un enfoque más focalizado, ‘saltando’ dentro de la clasificación para identificar comparables que coincidan estrechamente con el atractivo del inmueble a valorar. A medida que el rango se estrecha, el VE presta mayor atención a las similitudes entre los inmuebles, particularmente en aspectos como la calidad, el estado, la estética del diseño (incluidos elementos como la pintura, la carpintería, la cantería y las ventanas), la utilidad funcional y otras características relevantes

Es crucial comprender que los juicios del VE están fundamentalmente guiados por las clasificaciones establecidas por el mercado. Esto asegura que los elementos subjetivos del proceso estén anclados en las realidades del mercado, en procura de una valuación justa y exacta del inmueble a valorar.

12.5. Conclusión de valor del RCA

12.5.1. Consideraciones generales

Determinar una puntuación CQA es fundamental para establecer el valor residual estimado de un inmueble, ya que estos elementos están intrínsecamente vinculados. Si bien es posible evaluar un inmueble sin una CQA, esta métrica proporciona un contexto valioso al ilustrar la posición del inmueble en comparación con otros en la zona de mercado. Por ejemplo, una puntuación CQA de 3.2 indica que el inmueble tiene mayor atractivo que el 32 % de los inmuebles comparables en las cercanías. Si existen objeciones a la puntuación asignada, los objetores deben consultar las clasificaciones y justificar cualquier ajuste propuesto con base en evidencia fotográfica u otros datos sustanciales. Si el sistema de clasificación es robusto y el Valuation Engineer (VE) ha posicionado con exactitud el inmueble dentro de este marco, es poco probable que surjan discrepancias significativas.

En la fase de elaboración del informe, típicamente se seleccionan de 6 a 12 comparables de venta recientes y relevantes para respaldar la conclusión de la valuación. Se analizan las diferencias entre estos comparables y el inmueble a valorar, ajustándose cada variable en consecuencia. Estos ajustes se aplican luego a los precios de venta netos de los comparables para derivar los precios de venta ajustados. Como se demostró matemáticamente, estos precios ajustados deben converger con el precio de venta estimado del inmueble a valorar.

12.5.2. Revisión, auditoría, quejas

Una vez determinado el residuo estimado del inmueble a valorar, pueden calcularse los ajustes residuales para todos los comparables. Estos ajustes sirven como restricciones para desgloses manuales adicionales por parte del VE, adaptados a las variables CQA específicas que se consideren pertinentes. Es importante enfatizar que estos desgloses detallados son explicativos y no influyen en la valuación final, como se afirma en la Demostración II.

En cuanto a las quejas presentadas por usuarios no satisfechos con la conclusión de valor del avalúo, un valuator tradicional típicamente tiene que perder tiempo explicando por qué tal o cual inmueble con un precio más alto o más bajo no tiene el impacto en el valor que el usuario esperaba. Con el RCA, todas las ventas comparables posibles (usualmente 100-600) se evalúan típicamente exactamente al mismo valor que la conclusión de precio, por lo que se requiere

poco esfuerzo para mostrar los ajustes de algún inmueble arbitrario. Todo el proceso del RCA es automatizado y objetivo, excepto por la puntuación del residuo del inmueble a valorar. Y así, otra ventaja del RCA es que la crítica por imperfecciones en el juicio subjetivo del VE queda principalmente restringida a esa única área - y puede gestionarse de manera efectiva y eficiente. Además, suponiendo un R^2 de alrededor del 80 %, el impacto de un error de $\pm 10\%$ en la estimación del residuo del inmueble a valorar equivale a un error de $\pm 2\%$ en la estimación final del valor.

En muchos casos, el inmueble a valorar encaja tan perfectamente en la clasificación residual de los inmuebles, que puede esperarse que la exactitud esté dentro de $\pm 1\%$.

La suma del residuo estimado y el precio de venta del inmueble a valorar estimado por MARS produce la conclusión de valor final, conocida como la Conclusión de Valor del Residual Constraint Approach (RCA). Este proceso asegura un enfoque integral y transparente de la valuación inmobiliaria, fomentando la confianza en la exactitud y la equidad de la evaluación.

13. Conclusión

El Residual Constraint Approach es un marco con tres contribuciones sustantivas: una propiedad estructural anti-anclaje derivada de ajustar el modelo sobre el mercado completo en lugar de sobre un conjunto reducido de comparables (Demostración I); un resultado de invariancia que permite que la descomposición del residuo sirva como capa explicativa sin alterar la conclusión de valor (Demostración II y la observación que la sigue); y una regularidad empírica — la curva característica CQA-Residuo — que parece repetirse en distintas zonas de mercado con la forma característica exponencial-lineal-exponencial. Las limitaciones del marco son igualmente sustantivas: la descomposición del residuo es invariante pero no identificadora, el paso de asignación de CQA depende del juicio del VE, que aún no ha sido validado mediante un estudio formal de concordancia entre evaluadores, y las afirmaciones de exactitud están pendientes de una evaluación comparativa frente a alternativas.

Lo que este artículo presenta es un marco y un protocolo. Lo que aún no presenta, pero sí lo hace el programa de investigación más amplio actualmente en curso, es una cadena de procesamiento ejecutable y auditable que operacionaliza el protocolo en software con valores predeterminados defendibles. Esa cadena de procesamiento — el paquete earthUI en CRAN para la etapa de funciones base de MARS, el paquete glmnetUI para el modelo estructural regularizado sobre dichas bases, y el paquete mgcvUI para el suavizado espacial — es el tema de artículos complementarios en este número y constituye el puente del RCA-como-método al RCA-como-entregable. La brecha entre doctrina y práctica en la regulación de la valuación (Craytor, 2026, este número) aporta la razón por la que ese puente importa: los productos de la valuación se consumen económicamente como insumos de riesgo prospectivos, y un marco que hace que la opinión de valor presente sea auditable y reproducible es una condición previa para cualquier tratamiento honesto de ese uso prospectivo. El RCA es un movimiento dentro de ese proyecto más amplio. No es el proyecto completo, y este artículo no pretende que lo sea.

Referencias

- Ambrose, Brent W. et al. (2025). «Do Appraiser and Borrower Race Affect Mortgage Collateral Valuation?» En: *Review of Finance*. DOI: [10.1093/rof/rfaf046](https://doi.org/10.1093/rof/rfaf046).
- Craytor, William Bert (2026). «The Doctrine–Practice Gap in Real Estate Appraisal: A Structured Account of Functions, Boundaries, and Tensions». En: *Valuation Engineer Journal* 1.1.
- Friedman, Jerome H. (1991). «Multivariate Adaptive Regression Splines». En: *The Annals of Statistics* 19.1, págs. 1-141.
- Hastie, Trevor, Robert Tibshirani y Jerome H. Friedman (2017). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Second. Springer Series in Statistics. New York, NY: Springer. ISBN: 978-0-387-84857-0.
- James, Gareth et al. (2013). *An Introduction to Statistical Learning: With Applications in R*. Springer Texts in Statistics 103. New York: Springer. ISBN: 978-1-4614-7137-0.
- Krivorotov, George y Michael LaCour-Little (2020). «AVM versus Appraisal-Based Underwriting in Refinance Mortgages: The Trade-off Between Noise and Bias». En: *SSRN Electronic Journal*. DOI: [10.2139/ssrn.3603327](https://doi.org/10.2139/ssrn.3603327).
- Kuhn, Max y Kjell Johnson (2016). *Applied Predictive Modeling*. Corrected at 5th printing. New York: Springer. ISBN: 978-1-4614-6848-6.
- Milborrow, Stephen (2024a). *Earth: Multivariate Adaptive Regression Splines*. R Package. URL: <https://CRAN.R-project.org/package=earth>.
- (2024b). *Notes on the Earth Package*. Inf. téc. URL: <http://www.milbo.org/doc/earth-notes.pdf>.

- Milborrow, Stephen (2024c). *Variance Models in Earth*. Inf. téc. URL: <http://www.milbo.org/doc/earth-varmod.pdf>.
- Perry, Andre M., Jonathan Rothwell y David Harshbarger (2018). *The Devaluation of Assets in Black Neighborhoods*. Inf. téc. Brookings Institution.
- Perry, Andre M., Jonathan Rothwell, J. Williamson et al. (2024). *How Racial Bias in Appraisals Affects the Devaluation of Homes in Majority-Black Neighborhoods*. Inf. téc. Brookings Institution.
- Property Appraisal and Valuation Equity Task Force (2022). *Action Plan to Advance Property Appraisal and Valuation Equity*. Inf. téc. U.S. Department of Housing y Urban Development.
- Tzioumis, Konstantinos (2016). «Appraisers and Valuation Bias: An Empirical Analysis». En: *Real Estate Economics* 45.3, págs. 679-712. DOI: [10.1111/1540-6229.12133](https://doi.org/10.1111/1540-6229.12133).
- Williamson, Jake y Mark Palim (2022). *Appraising the Appraisal: A Closer Look at Divergent Appraisal Values for Black and White Borrowers Refinancing Their Home*. Inf. téc. Fannie Mae.
- Zhang, Wengang (2020). *MARS Applications in Geotechnical Engineering Systems: Multi-Dimension with Big Data*. Singapore: Springer. ISBN: 978-981-13-7421-0.