

Reglas de Asociación

Con el objetivo de facilitar el análisis de reglas de asociación, se realizó una transformación de las variables de la base de datos. Las variables numéricas **precio** y **superficie** se agruparon en cuartiles, generando variables categóricas como “precio_bajo” o “superficie_alta”. Asimismo, se transformaron variables booleanas en etiquetas comprensibles (por ejemplo, **tieneAscensor** = 1 se convierte en "**con_ascensor**"), eliminando redundancias y garantizando la consistencia semántica de los datos.

Una vez procesados los datos, se aplicó el algoritmo **Apriori** con un umbral mínimo de soporte del 1% y confianza del 50% obteniendo un total de 12745 reglas. Posteriormente se han eliminado reglas redundantes, aquellas que no aportan información adicional respecto a reglas más generales con igual o mejor confianza. Esto permite conservar las reglas más significativas. Las reglas maximales no se han priorizado, ya que suelen perder especificidad, dificultando su aplicación práctica. Tras eliminar las redundantes nos quedamos con un total de 2408 reglas. Por último, filtramos las reglas con los siguientes umbrales quedándonos con las 21 reglas más relevantes.

- Soporte > 0.015 | Confianza > 0.7 | Lift > 3.5

Estas condiciones permiten identificar patrones **robustos y estadísticamente relevantes** que relacionan ciertas configuraciones de un inmueble con su probabilidad de pertenecer a un rango de precio específico. A continuación se muestran las **cinco reglas más destacadas**, ordenadas por su *lift*. Es importante destacar que las 21 reglas comparten el consecuente de **Precio Alto**.

Table 1: Reglas de asociación más relevantes (ordenadas por lift)

	rules	support	confidence	coverage	lift	count
1001	{rooms_4+_habitaciones,tieneAscensor_con_ascensor,bathrooms_3+_banios} => {priceAmount_precio_alto}	0.017	0.970	0.018	3.879	32
8593	{surface_surface_alto,tieneAscensor_con_ascensor,tieneAireAcondicionado_con_aire,bathrooms_3+_banios,tieneCalefaccion_con_calefaccion} => {priceAmount_precio_alto}	0.017	0.970	0.018	3.879	32
4075	{surface_surface_alto,tieneAscensor_con_ascensor,tieneAireAcondicionado_con_aire,bathrooms_3+_banios} => {priceAmount_precio_alto}	0.022	0.952	0.023	3.810	40
4063	{surface_surface_alto,tieneAscensor_con_ascensor,bathrooms_3+_banios,tieneCalefaccion_con_calefaccion} => {priceAmount_precio_alto}	0.018	0.943	0.019	3.771	33
4284	{tieneAscensor_con_ascensor,tieneAireAcondicionado_con_aire,bathrooms_3+_banios,tieneCalefaccion_con_calefaccion} => {priceAmount_precio_alto}	0.018	0.943	0.019	3.771	33

Una vez identificadas dichas reglas y analizando que están asociadas a precios altos, exploramos ahora aquellas cuya consecuencia es priceAmount = precio_bajo. Este tipo de reglas resulta especialmente útil para detectar inmuebles infravalorados o con condiciones objetivas que los hacen significativamente más asequibles. Para ello, filtramos las reglas con:

- Soporte > 0.0107 | Confianza > 0.8 | Lift > 2.7

Table 2: Reglas de asociación más relevantes asociadas a Precios Bajos (ordenadas por lift)

	rules	support	confidence	coverage	lift	count
4271	{tieneAscensor_sin_ascensor,rooms_1_habitacion,surface_surface_bajo,tieneAireAcondicionado_sin_aire} => {priceAmount_precio_bajo}	0.011	0.840	0.013	3.360	21
8705	{tieneTrastero_sin_trastero,tieneCalefaccion_sin_calefaccion,rooms_1_habitacion,surface_surface_bajo,tieneAireAcondicionado_sin_aire} => {priceAmount_precio_bajo}	0.012	0.821	0.015	3.286	23
4284	{tieneCalefaccion_sin_calefaccion,rooms_1_habitacion,surface_surface_bajo,tieneAireAcondicionado_sin_aire} => {priceAmount_precio_bajo}	0.013	0.806	0.017	3.226	25

A continuación, se han empleado dos tipos de gráficos para explorar visualmente las reglas de asociación más representativas: un **diagrama de dispersión (scatter plot)** para analizar la relación entre soporte y confianza, y un **gráfico de coordenadas paralelas** para examinar la estructura interna de las reglas en términos de los atributos que las componen.

No se utilizaron representaciones como el grouped matrix ni la matriz de calor, ya que estas técnicas son más adecuadas cuando las reglas implican combinaciones simples entre pocos ítems. En este caso, las reglas seleccionadas están formadas por múltiples atributos simultáneos, lo que genera una elevada dimensionalidad y dificulta la interpretación gráfica en formatos matriciales. Por ello, se optó por visualizaciones más eficaces para este tipo de complejidad estructural.

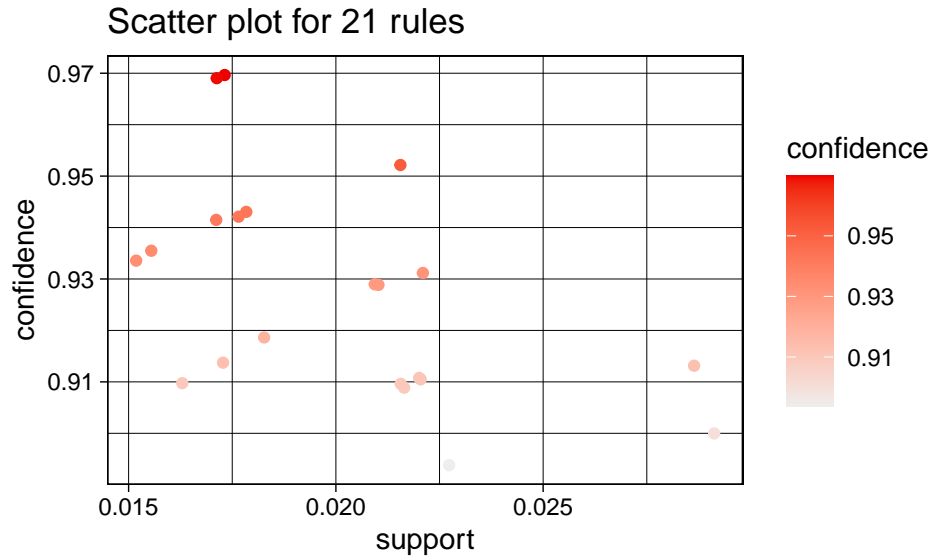


Figure 1: Distribución de Reglas

El gráfico de dispersión muestra la relación entre el soporte y la confianza de las 21 reglas de asociación seleccionadas, con una escala de color que representa el nivel de confianza. Comprobamos que todas las reglas tienen un soporte comprendido entre 0.015 y 0.025, lo cual indica que aunque no son extremadamente frecuentes, sí aparecen con suficiente consistencia en la base de datos. En términos de confianza, todas superan el umbral mínimo de 0.7, alcanzando valores próximos al 0.97. Esto indica que cuando se cumplen las condiciones del antecedente, la probabilidad de que el precio sea alto es muy elevada. Las reglas con mayor confianza tienden a tener un soporte ligeramente menor, lo que sugiere que, si bien son altamente fiables, aplican a un subconjunto más específico del mercado inmobiliario.

Parallel coordinates plot for 21 rules

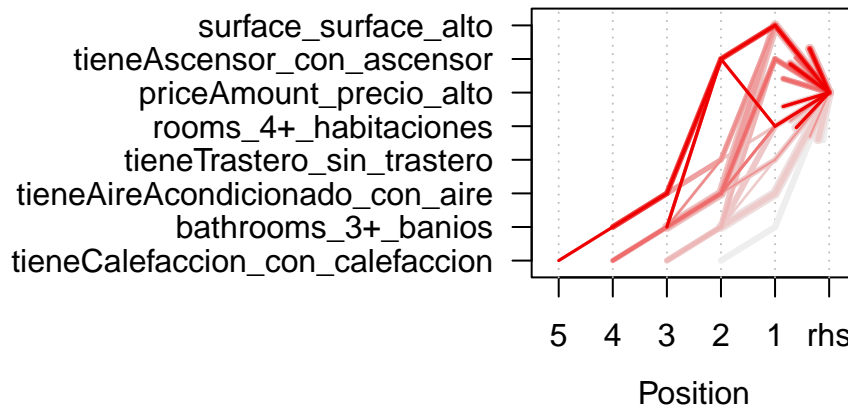


Figure 2: Gráfico de Coordenadas Paralelas

El gráfico de coordenadas paralelas representa visualmente la estructura interna de las reglas más representativas. Cada línea corresponde a una regla individual, y su recorrido conecta los distintos atributos que forman el antecedente, finalizando en el consecuente (`priceAmount_precio_alto`).

Se aprecia una clara convergencia de reglas hacia ciertos atributos comunes, como:

- Superficie elevada del inmueble (`surface_surface_alto`)
- Existencia de ascensor (`tieneAscensor_con_ascensor`)
- Más de 3 baños (`bathrooms_3+_banios`)
- Existencia de calefacción (`tieneCalefaccion_con_calefaccion`)

Estas características aparecen reiteradamente como predictores del precio alto, lo que refuerza su relevancia dentro del conjunto de datos analizado.

El análisis de reglas de asociación ha revelado patrones sólidos entre las características estructurales de los inmuebles y su rango de precio. En particular, propiedades con gran superficie, múltiples baños y habitaciones, ascensor, calefacción y aire acondicionado muestran una alta probabilidad de pertenecer al segmento de precio alto.

En contraste, los inmuebles con superficie reducida, una sola habitación y ausencia de comodidades tienden a asociarse con precios bajos, lo que puede indicar oportunidades de inversión o infravaloración.

Estos resultados proporcionan una base objetiva y cuantitativa para la valoración rápida de propiedades, con aplicaciones directas en la toma de decisiones de compradores, inversores y plataformas de tasación automatizada.