

A1_U1

Juan José Restrepo Rosero

2025-02-09

Contents

Análisis exploratorio de datos	1
1. Análisis de Componentes Principales (PCA):	11
2. Análisis de Conglomerados (Clustering):	16
3. Análisis de Correspondencia:	17

Informe Evaluación de la oferta inmobiliaria urbana

Presentado por Juan José Restrepo Rosero

Problema

Una empresa inmobiliaria líder en una gran ciudad está buscando comprender en profundidad el mercado de viviendas urbanas para tomar decisiones estratégicas más informadas. La empresa posee una base de datos extensa que contiene información detallada sobre diversas propiedades residenciales disponibles en el mercado. Se requiere realizar un análisis holístico de estos datos para identificar patrones, relaciones y segmentaciones relevantes que permitan mejorar la toma de decisiones en cuanto a la compra, venta y valoración de propiedades.

Retos:

El reto principal consisten en realizar un análisis integral y multidimensional de la base de datos para obtener una comprensión del mercado inmobiliario urbano. Se requiere aplicar diversas técnicas de análisis de datos, incluyendo

Análisis de Componentes Principales: Reducir la dimensionalidad del conjunto de datos y visualizar la estructura de las variables en componentes principales para identificar características clave que influyen en la variación de precios y oferta del mercado.

Análisis de Conglomerados: Agrupar las propiedades residenciales en segmentos homogéneos con características similares para entender las dinámicas de las ofertas específicas en diferentes partes de la ciudad y en diferentes estratos socio económicos.

Análisis de Correspondencia : Examinar la relación entre las variables categóricas (tipo de vivienda, zona y barrio), para identificar patrones de comportamiento de la oferta en mercado inmobiliario.

Visualización de resultados: Presentar gráficos, mapas y otros recursos visuales para comunicar los hallazgos de manera clara y efectiva a la dirección de la empresa.

Análisis exploratorio de datos

En esta etapa inicial, como se presenta durante el momento de *Acción* en Brightspace, se procede a explorar el conjunto de datos para obtener una visión compacta de su dimensión y atributos. Es importante resaltar

que la base de datos de vivienda consta de 8,322 filas y 13 columnas, información clave para identificar las características particulares de los datos, tal como se analiza en la unidad de análisis multivariado.

```
## spc_tbl_ [8,322 x 13] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ id      : num [1:8322] 1147 1169 1350 5992 1212 ...
## $ zona    : chr [1:8322] "Zona Oriente" "Zona Oriente" "Zona Oriente" "Zona Sur" ...
## $ piso    : chr [1:8322] NA NA NA "02" ...
## $ estrato  : num [1:8322] 3 3 3 4 5 5 4 5 5 5 ...
## $ preciom  : num [1:8322] 250 320 350 400 260 240 220 310 320 780 ...
## $ areaconst : num [1:8322] 70 120 220 280 90 87 52 137 150 380 ...
## $ parqueaderos: num [1:8322] 1 1 2 3 1 1 2 2 2 2 ...
## $ banios   : num [1:8322] 3 2 2 5 2 3 2 3 4 3 ...
## $ habitaciones: num [1:8322] 6 3 4 3 3 3 3 4 6 3 ...
## $ tipo     : chr [1:8322] "Casa" "Casa" "Casa" "Casa" ...
## $ barrio   : chr [1:8322] "20 de julio" "20 de julio" "20 de julio" "3 de julio" ...
## $ longitud : num [1:8322] -76.5 -76.5 -76.5 -76.5 -76.5 ...
## $ latitud  : num [1:8322] 3.43 3.43 3.44 3.44 3.46 ...
## - attr(*, "spec")=
## .. cols(
## ..   id = col_double(),
## ..   zona = col_character(),
## ..   piso = col_character(),
## ..   estrato = col_double(),
## ..   preciom = col_double(),
## ..   areaconst = col_double(),
## ..   parqueaderos = col_double(),
## ..   banios = col_double(),
## ..   habitaciones = col_double(),
## ..   tipo = col_character(),
## ..   barrio = col_character(),
## ..   longitud = col_double(),
## ..   latitud = col_double()
## .. )
## - attr(*, "problems")=<externalptr>
```

En esta sección, se observa que, según la muestra de datos, el campo “piso” muestra valores ausentes, los cuales parecen estar relacionados con el tipo de vivienda denominado “casa”. No obstante, esta primera inferencia no es completamente precisa, ya que existen registros del mismo tipo de vivienda que sí contienen dicha información.

```
options(tibble.width = Inf)
kable(head(vivienda))
```

id	zona	piso	estrato	preciom	areaconst	parqueaderos	banios	habitaciones	tipo	barrio	longitud	latitud
1147	Zona Oriente	NA	3	250	70	1	3	6	Casa	20 de julio	- 76.51168	3.43382
1169	Zona Oriente	NA	3	320	120	1	2	3	Casa	20 de julio	- 76.51237	3.43369
1350	Zona Oriente	NA	3	350	220	2	2	4	Casa	20 de julio	- 76.51537	3.43566
5992	Zona Sur	02	4	400	280	3	5	3	Casa	3 de julio	- 76.54000	3.43500

id	zona	piso	estrato	preciom	areaconst	parqueaderos	banios	habitaciones	tipo	barrio	longitud	latitud
1212	Zona Norte	01	5	260	90	1	2	3	Apartamento	atopi	- 76.51350	3.45891
1724	Zona Norte	01	5	240	87	1	3	3	Apartamento	atopi	- 76.51700	3.36971

En esta sección, se busca obtener un resumen estadístico básico del conjunto de datos. En este resumen, se puede identificar que las columnas “zona”, “piso”, “tipo” y “barrio” corresponden a variables categóricas.

id	zona	piso	estrato	preciom	areaconst	parqueaderos	banios	habitaciones	tipo	barrio	longitud	latitud
Min. : 1	Length:8122	Length:8322	Min. :3.000	Min. :58.0	Min. :30.0	Min. :1.000	Min. :0.000	Min. :0.000	Length:8122	Length:8322	Min. :- 76.59	Min. :3.333
1st Qu.:2080	Class :character	Class :character	1st Qu.:4.000	1st Qu.:220.0	1st Qu.:80.0	1st Qu.:1.000	1st Qu.:2.000	1st Qu.:3.000	Class :character	Class :character	1st Qu.:76.54	1st Qu.:3.381
Median :4160	Mode :character	Mode :character	Median :5.000	Median :330.0	Median :123.0	Median :2.000	Median :3.000	Median :3.000	Mode :character	Mode :character	Median :- 76.53	Median :3.416
Mean :4160	NA	NA	Mean :4.634	Mean :433.9	Mean :174.9	Mean :1.835	Mean :3.111	Mean :3.605	NA	NA	Mean :- 76.53	Mean :3.418
3rd Qu.:6240	NA	NA	3rd Qu.:5.000	3rd Qu.:540.0	3rd Qu.:229.0	3rd Qu.:2.000	3rd Qu.:4.000	3rd Qu.:4.000	NA	NA	3rd Qu.:76.52	3rd Qu.:3.452
Max. :8319	NA	NA	Max. :6.000	Max. :1999.0	Max. :1745.0	Max. :10.000	Max. :10.000	Max. :10.000	NA	NA	Max. :- 76.46	Max. :3.498
NA's :3	NA	NA	NA's :3	NA's :2	NA's :3	NA's :1605	NA's :3	NA's :3	NA	NA	NA's :3	NA's :3

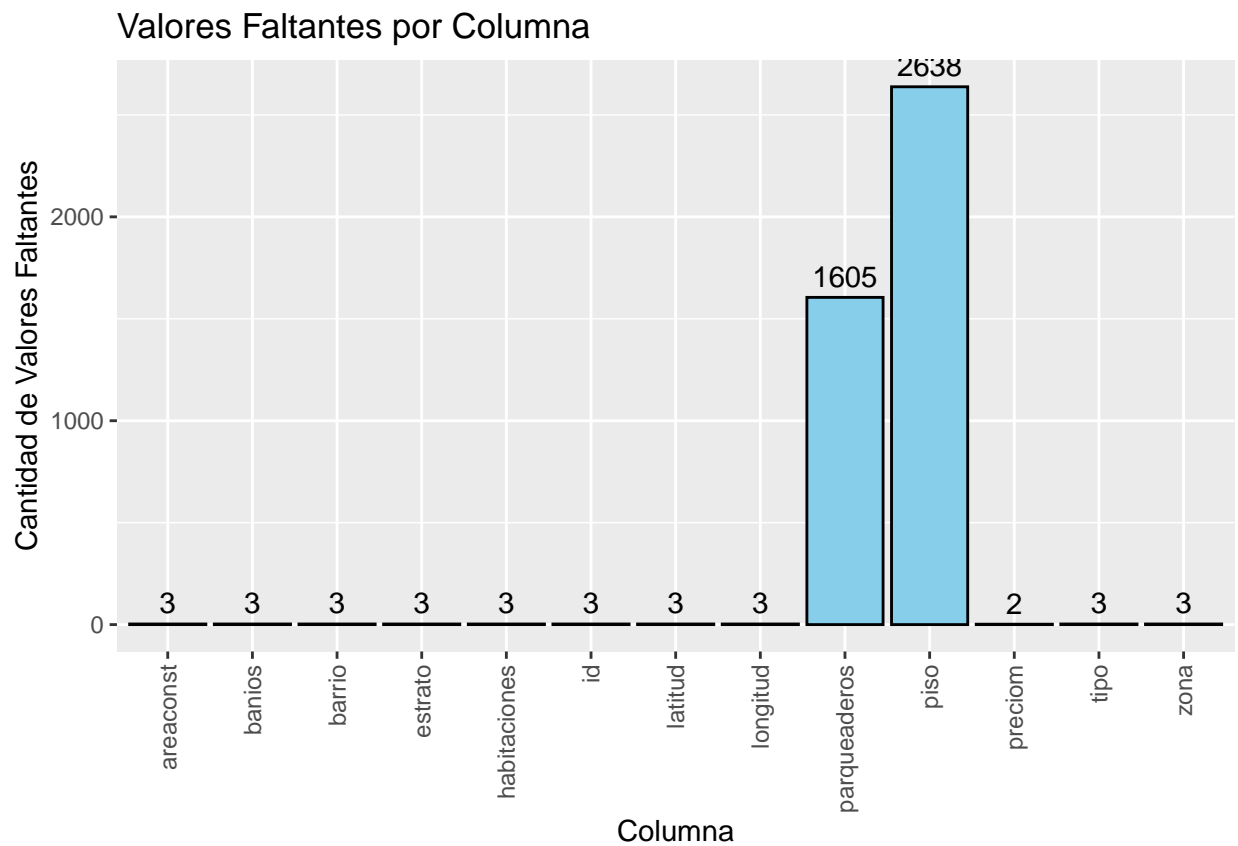
A partir de este análisis, se identifican los valores faltantes o ausentes, y se procede a realizar un enfoque detallado para evidenciar gráficamente su impacto. En la tabla anterior, el resumen estadístico mostraba un sesgo, indicando que el campo “parqueadero” tenía el mayor volumen de datos ausentes, con un total de 1605. Sin embargo, tras revisar los pasos previos, se detectó que la columna “piso” no presentaba valores ausentes, sino que se había rellenado con la cadena de texto “NA”. Por lo tanto, en los pasos siguientes se buscará verificar y corregir esta información.

	Columna	Valores_Faltantes
id	id	3
zona	zona	3
piso	piso	2638
estrato	estrato	3
preciom	preciom	2
areaconst	areaconst	3
parqueaderos	parqueaderos	1605
banios	banios	3
habitaciones	habitaciones	3
tipo	tipo	3
barrio	barrio	3

	Columna	Valores_Faltantes
	longitud	3
	latitud	3

Dado que todas las columnas del conjunto de datos contienen valores faltantes, se procede a representar gráficamente la proporción de estos valores ausentes en cada una de las columnas.

```
## Warning in geom_text(aes(label = Valores_Faltantes), vjust = -0.5, linewidth =
## 3): Ignoring unknown parameters: 'linewidth'
```

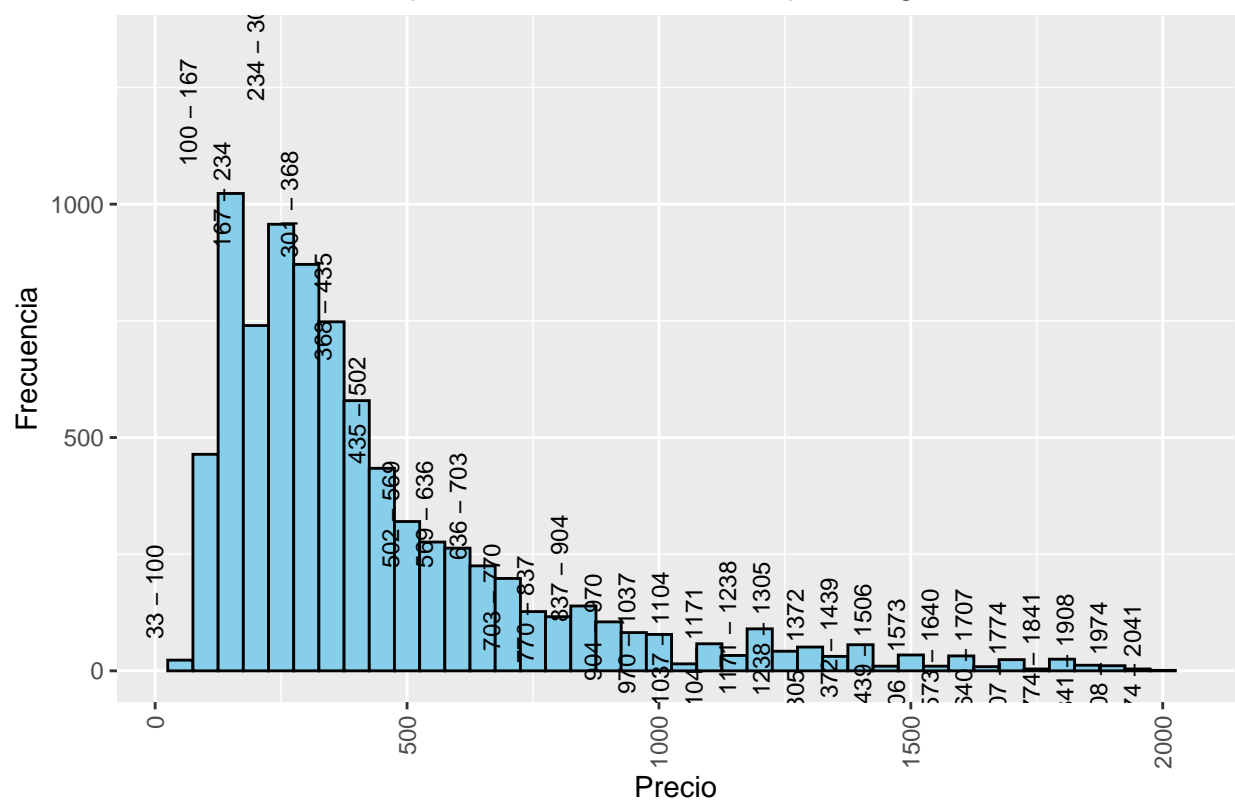


A continuación, se presentan, según mi criterio, las principales visualizaciones relacionadas con el negocio inmobiliario:

En el primer gráfico, se observa una alta concentración de registros en los rangos de precios 167-234 y 301-368. Para poder cuantificar adecuadamente estos inmuebles, será necesario excluir los valores “NA” durante el preprocesamiento de los datos.

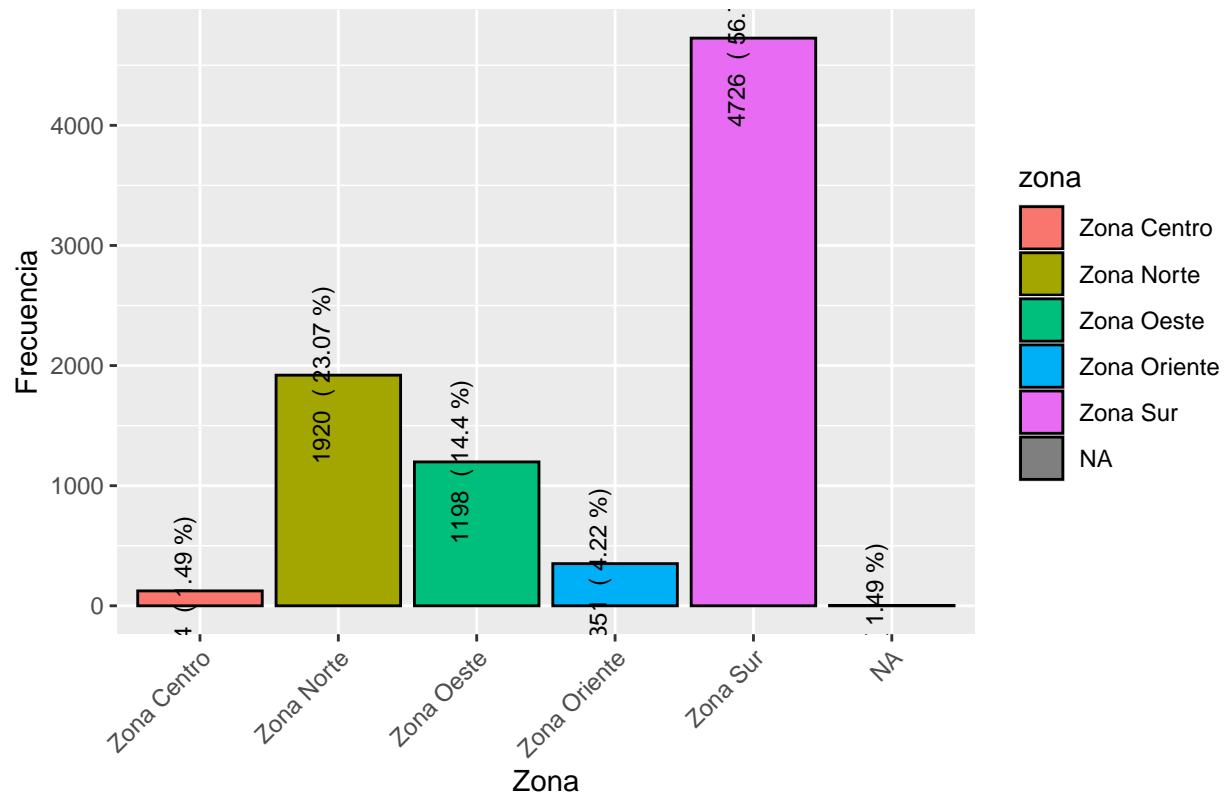
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

Distribucion de los precios de las viviendas por rangos



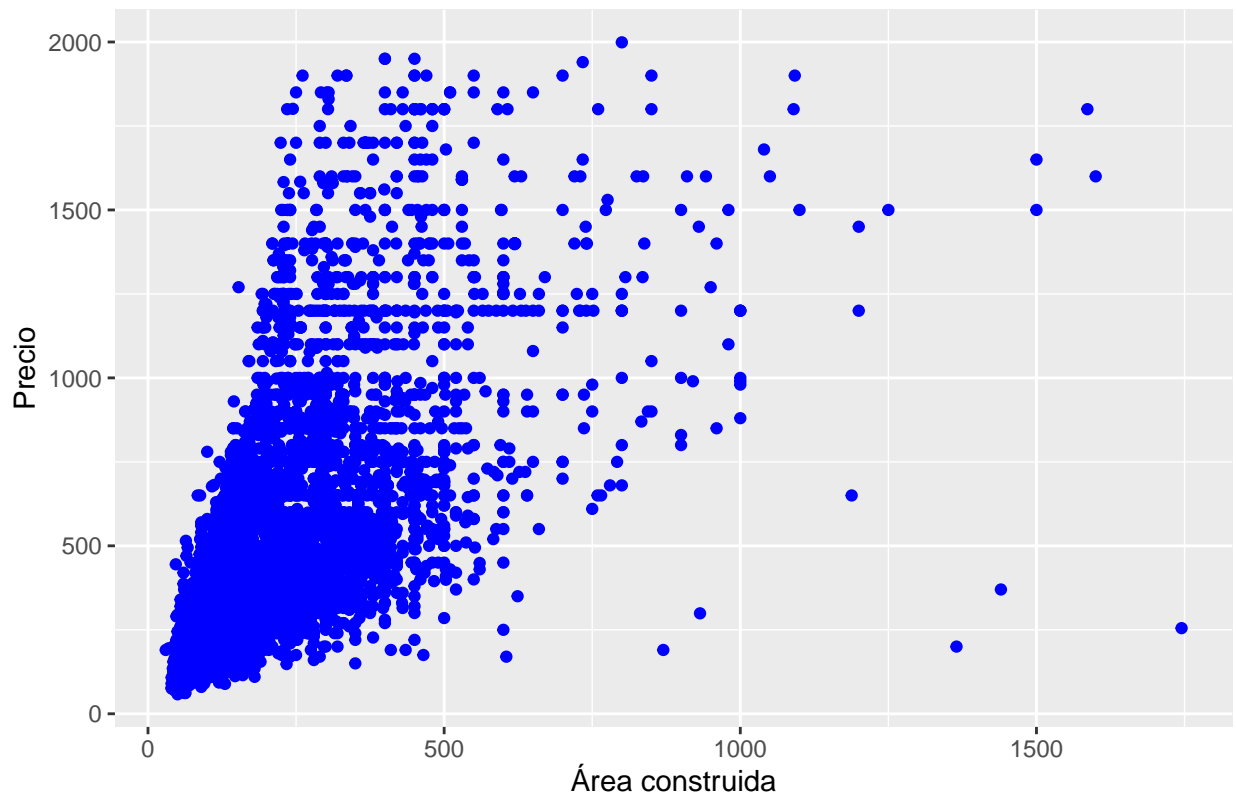
En el análisis anterior, es posible observar los datos ácidos, es decir, que se puede notar que las zonas más representativas en cuanto a registros de viviendas son la zona sur, con una participación del 56,8%, y la zona norte, con una participación del 20,1%. Juntas, estas zonas representan un total del 76,9%, lo que indica que la acción comercial más fuerte se concentra en estas áreas estratégicas de operación.

Distribución y participación de las viviendas por zona

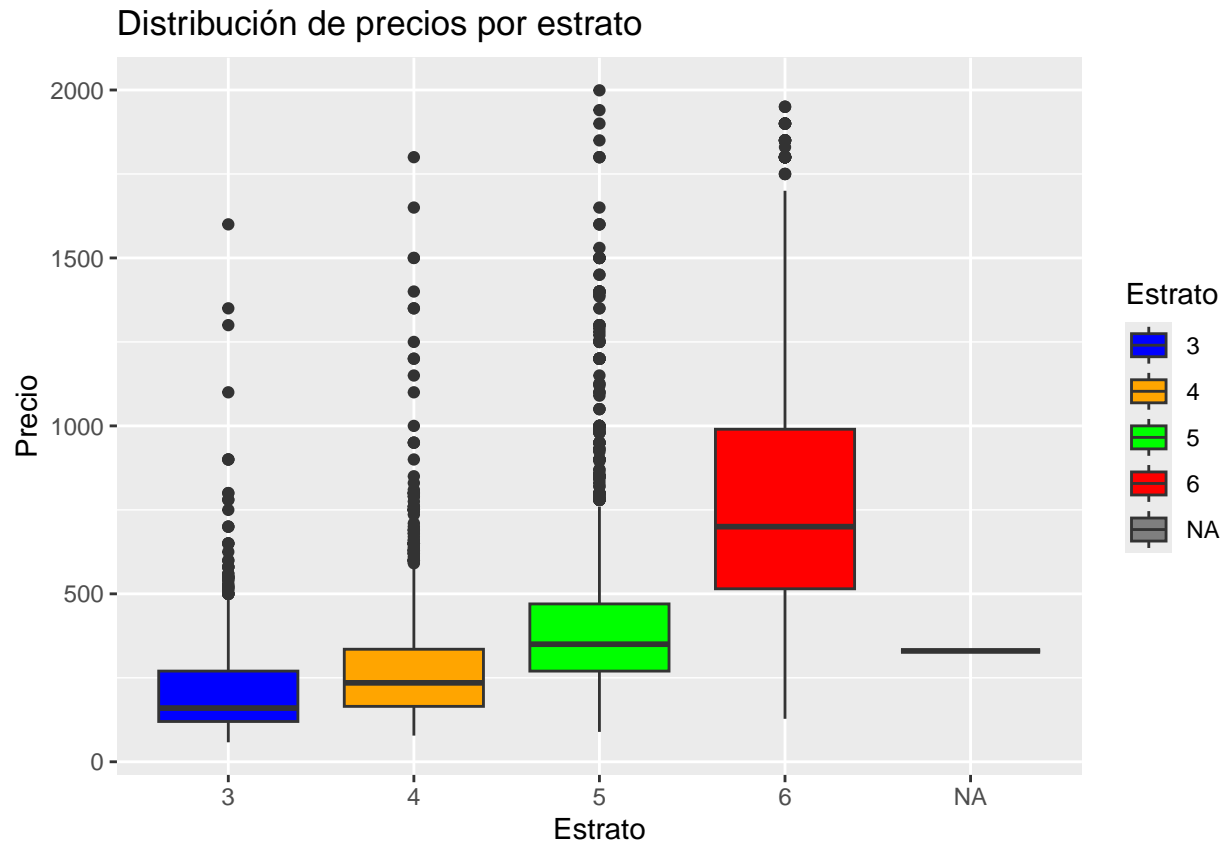


En este gráfico, se puede apreciar que la concentración de puntos cerca de la coordenada (250, 500) sugiere que existe una gran cantidad de viviendas con un área construida alrededor de las 250 unidades y un precio cercano a las 500 unidades. Este patrón podría indicar una tendencia común en el mercado inmobiliario, donde las viviendas con características similares (en términos de área) tienden a tener precios promedio establecidos de manera proporcional. Esto podría reflejar una relación entre el área construida y el precio, que podría ser un factor clave en la determinación del valor de estas propiedades.

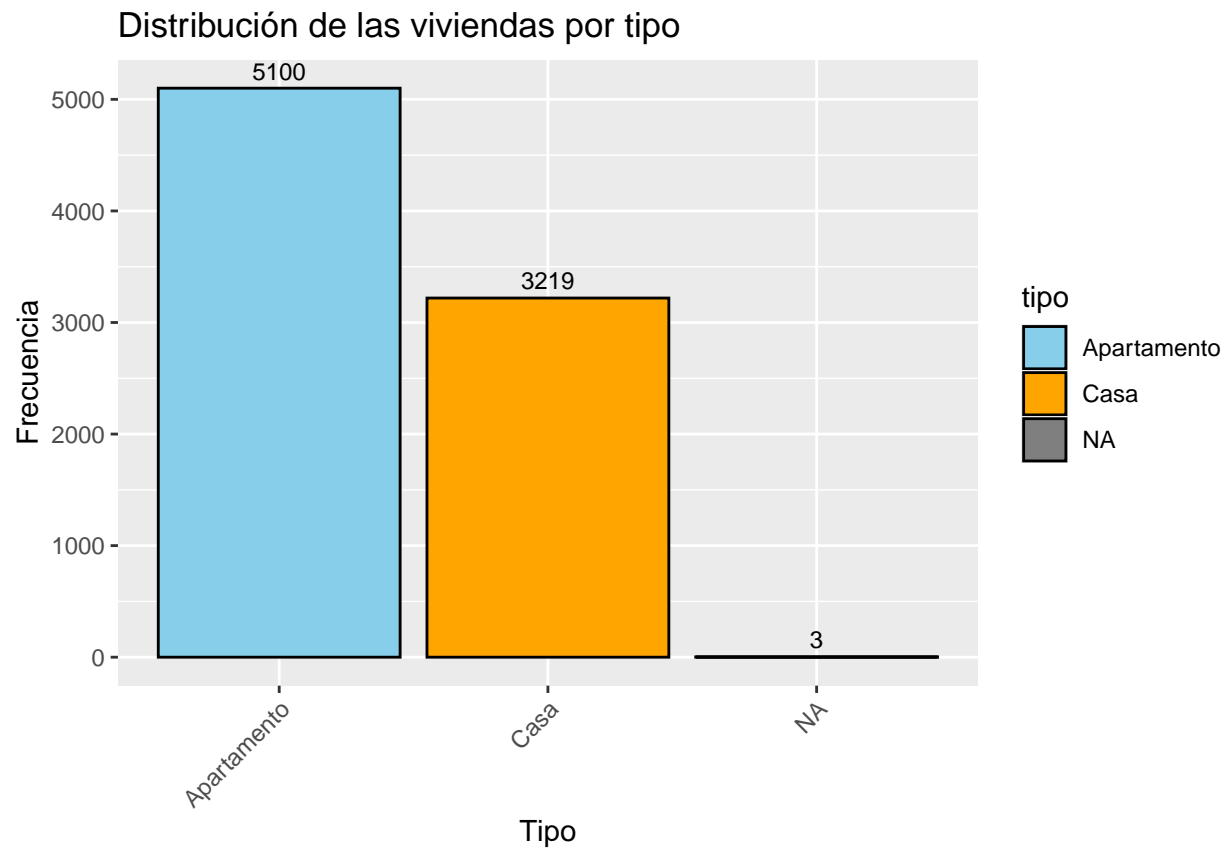
Relación entre el área construida y el precio



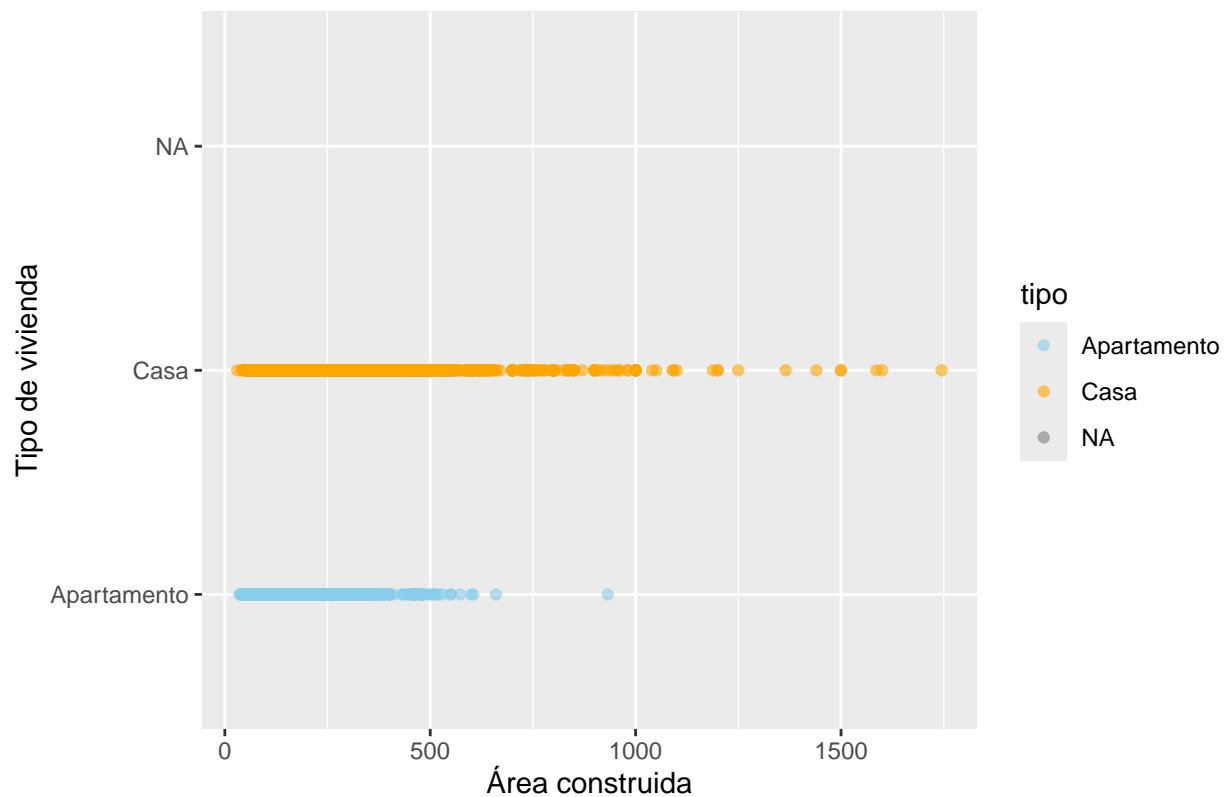
En este análisis, se observa que, en general, las viviendas clasificadas en el estrato 6 tienen precios significativamente más altos en comparación con las de otras categorías de estratos. Por otro lado, las viviendas del estrato 5 presentan precios en un rango más bajo en comparación con el estrato 6, pero aún así, sus precios son relativamente más altos que los de los estratos inferiores. Esto sugiere que el estrato 6 es el segmento más caro del mercado, seguido de cerca por el estrato 5, mientras que los estratos más bajos reflejan precios más accesibles.



En los siguientes gráficos se muestra la distribución de tipos de vivienda. En el primero, se observa que, aunque hay más apartamentos que casas en el conjunto de datos, el segundo gráfico revela que, al menos por tipo de vivienda, las casas tienen una mayor área construida en comparación con los apartamentos. Esto sugiere que, aunque la cantidad de apartamentos es mayor, las casas, en promedio, tienden a ser más grandes en términos de superficie construida.



Relación entre el área construida y el tipo de vivienda



Preprocesamiento de datos:

En base al análisis exploratorio anterior, se procede a eliminar las columnas ID, longitud y latitud, debido a que se considera que no aportan valor significativo al análisis del informe. Además, se realizará el tratamiento correspondiente a los valores faltantes o ausentes, que, como se observó durante la exploración de los datos, impidieron la evaluación de la correlación entre las variables. Estos valores serán gestionados adecuadamente para asegurar que no interfieran con los análisis posteriores, permitiendo un estudio más preciso y fiable.

```
##      zona      piso      estrato      preciom
## Length:8322      Length:8322      Min.   :3.000      Min.   : 58.0
## Class :character      Class :character      1st Qu.:4.000      1st Qu.: 220.0
## Mode  :character      Mode  :character      Median :5.000      Median : 330.0
##                                     Mean   :4.634      Mean   : 433.9
##                                     3rd Qu.:5.000      3rd Qu.: 540.0
##                                     Max.   :6.000      Max.   :1999.0
##                                     NA's   :3         NA's   :2
##      areaconst      parqueaderos      banios      habitaciones
## Min.   : 30.0      Min.   : 1.000      Min.   : 0.000      Min.   : 0.000
## 1st Qu.: 80.0      1st Qu.: 1.000      1st Qu.: 2.000      1st Qu.: 3.000
## Median :123.0      Median : 2.000      Median : 3.000      Median : 3.000
## Mean   :174.9      Mean   : 1.835      Mean   : 3.111      Mean   : 3.605
## 3rd Qu.:229.0      3rd Qu.: 2.000      3rd Qu.: 4.000      3rd Qu.: 4.000
## Max.   :1745.0      Max.   :10.000      Max.   :10.000      Max.   :10.000
## NA's   :3         NA's   :1605      NA's   :3         NA's   :3
##      tipo      barrio
## Length:8322      Length:8322
## Class :character      Class :character
```

```
## Mode :character Mode :character
##
##
##
##

## tibble [4,808 x 10] (S3: tbl_df/tbl/data.frame)
## $ zona      : chr [1:4808] "Zona Sur" "Zona Norte" "Zona Norte" "Zona Norte" ...
## $ piso      : chr [1:4808] "02" "01" "01" "01" ...
## $ estrato    : num [1:4808] 4 5 5 4 5 5 5 4 5 6 ...
## $ preciom    : num [1:4808] 400 260 240 220 310 320 780 625 750 520 ...
## $ areaconst  : num [1:4808] 280 90 87 52 137 150 380 355 237 98 ...
## $ parqueaderos: num [1:4808] 3 1 1 2 2 2 2 3 2 2 ...
## $ banios     : num [1:4808] 5 2 3 2 3 4 3 5 6 2 ...
## $ habitaciones: num [1:4808] 3 3 3 3 4 6 3 5 6 2 ...
## $ tipo       : chr [1:4808] "Casa" "Apartamento" "Apartamento" "Apartamento" ...
## $ barrio     : chr [1:4808] "3 de julio" "acopi" "acopi" "acopi" ...
```

1. Análisis de Componentes Principales (PCA):

A continuación se presentan los valores de la desviación estándar, los cuales reflejan el grado de dispersión de los datos que cada componente principal es capaz de explicar. Para simplificar este procedimiento, se seleccionaron únicamente las variables numéricas. En este contexto, la proporción de varianza y la proporción acumulada de varianza indican, respectivamente, el porcentaje de variabilidad total en los datos explicado por cada componente principal y por las componentes anteriores.

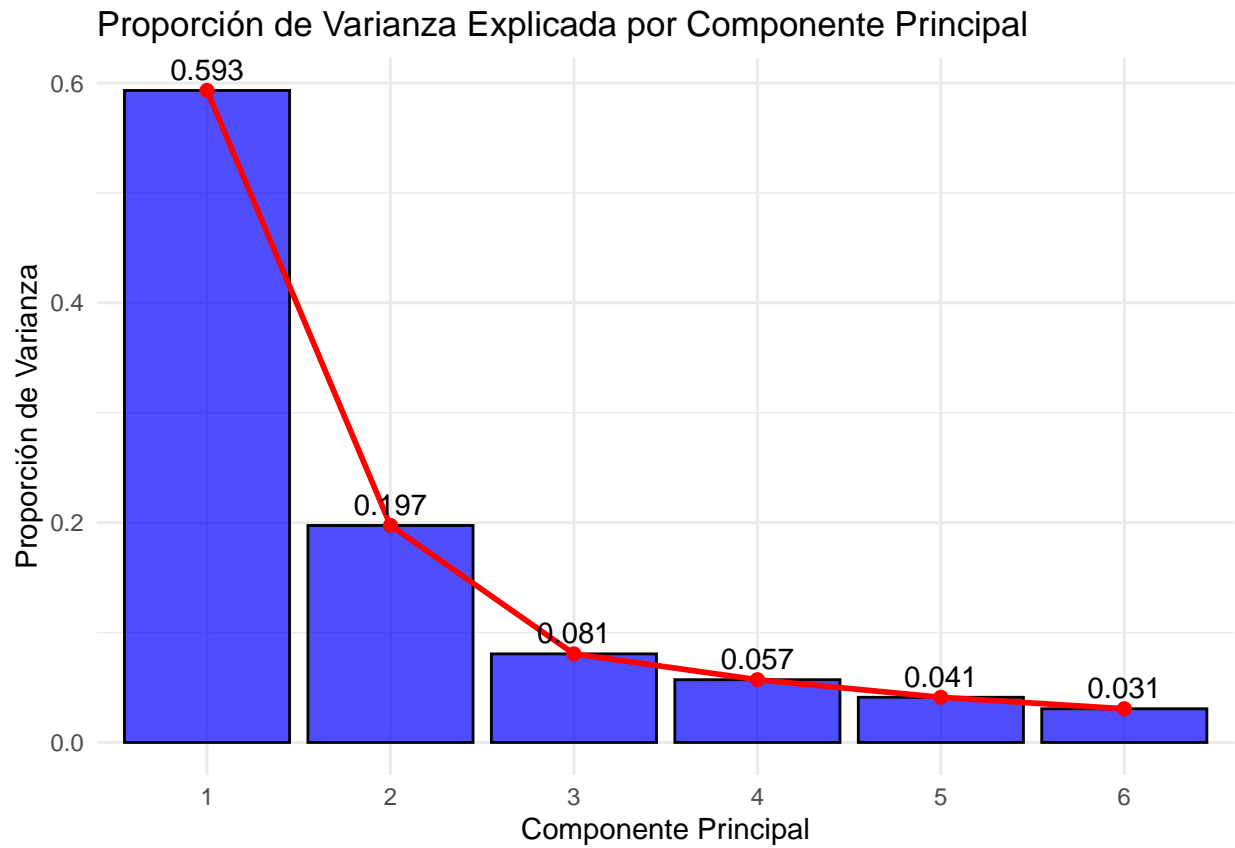
De acuerdo con estos cálculos:

- La *primera componente principal (PC1)* explica aproximadamente el 59.32% de la varianza total.
- Las *dos primeras componentes principales (PC1 y PC2)* juntas explican aproximadamente el 79.06% de la varianza total.
- Las *tres primeras componentes principales (PC1, PC2 y PC3)* explican aproximadamente el 87.12% de la varianza total.
- Las *cuatro primeras componentes principales (PC1, PC2, PC3 y PC4)* explican aproximadamente el 92.83% de la varianza total.
- Las *cinco primeras componentes principales (PC1, PC2, PC3, PC4 y PC5)* explican aproximadamente el 96.94% de la varianza total.
- Finalmente, *todas las seis componentes principales* explican el 100% de la varianza total, lo cual es esperado ya que se están utilizando todas las variables en el análisis.

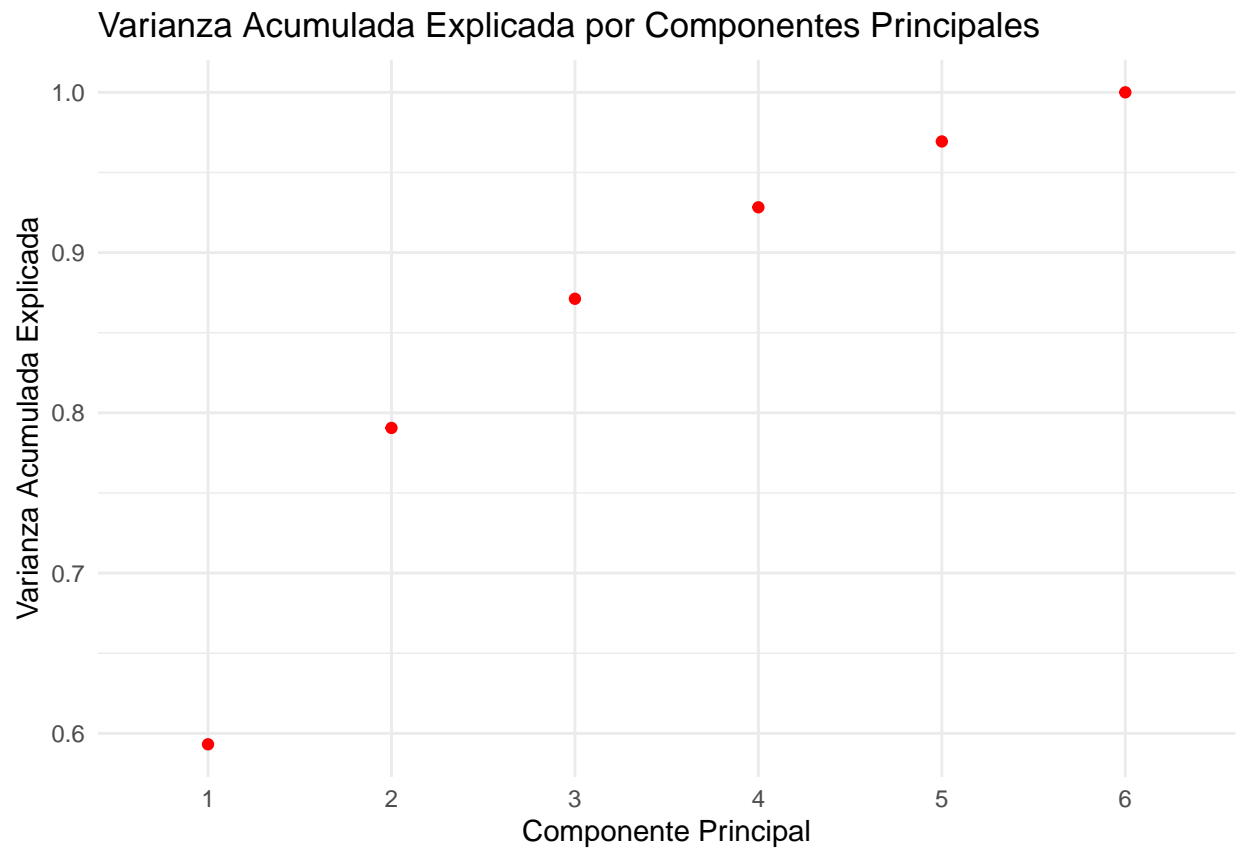
```
## Importance of components:
##           PC1      PC2      PC3      PC4      PC5      PC6
## Standard deviation  1.8866 1.0884 0.69530 0.58528 0.49647 0.42886
## Proportion of Variance 0.5932 0.1974 0.08057 0.05709 0.04108 0.03065
## Cumulative Proportion 0.5932 0.7906 0.87117 0.92827 0.96935 1.00000
```

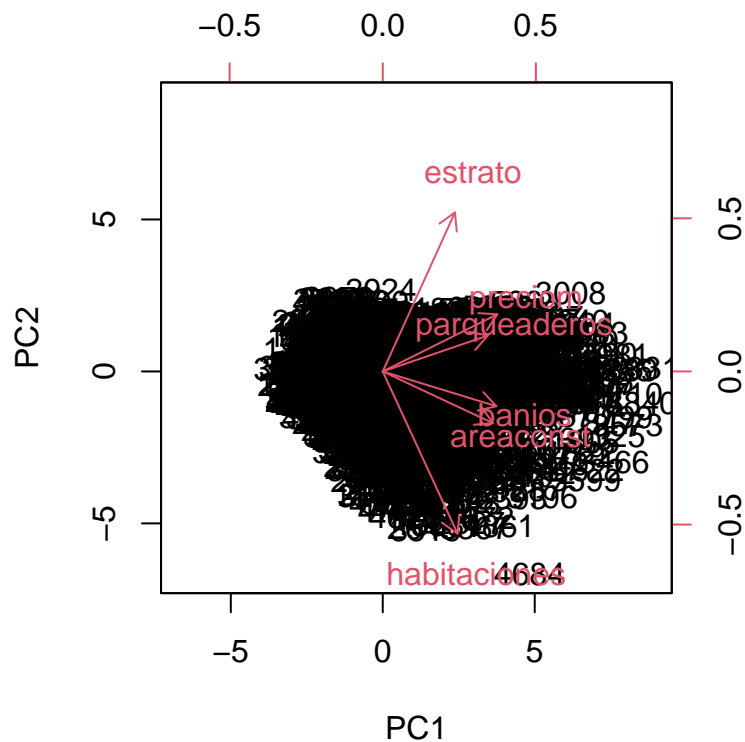
En este orden de ideas, se puede apreciar que el análisis de componentes principales (PCA) muestra un total de seis componentes, pero se observa que la mayor parte de la varianza es explicada por las dos primeras componentes principales (PC1 y PC2), que juntas capturan el 79.06% de la varianza total. Este hallazgo sugiere que, para simplificar el análisis, bastaría con considerar solo estas dos primeras componentes, ya que representan la mayor parte de la variabilidad presente en los datos.

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.  
## i Please use 'linewidth' instead.  
## This warning is displayed once every 8 hours.  
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was  
## generated.
```



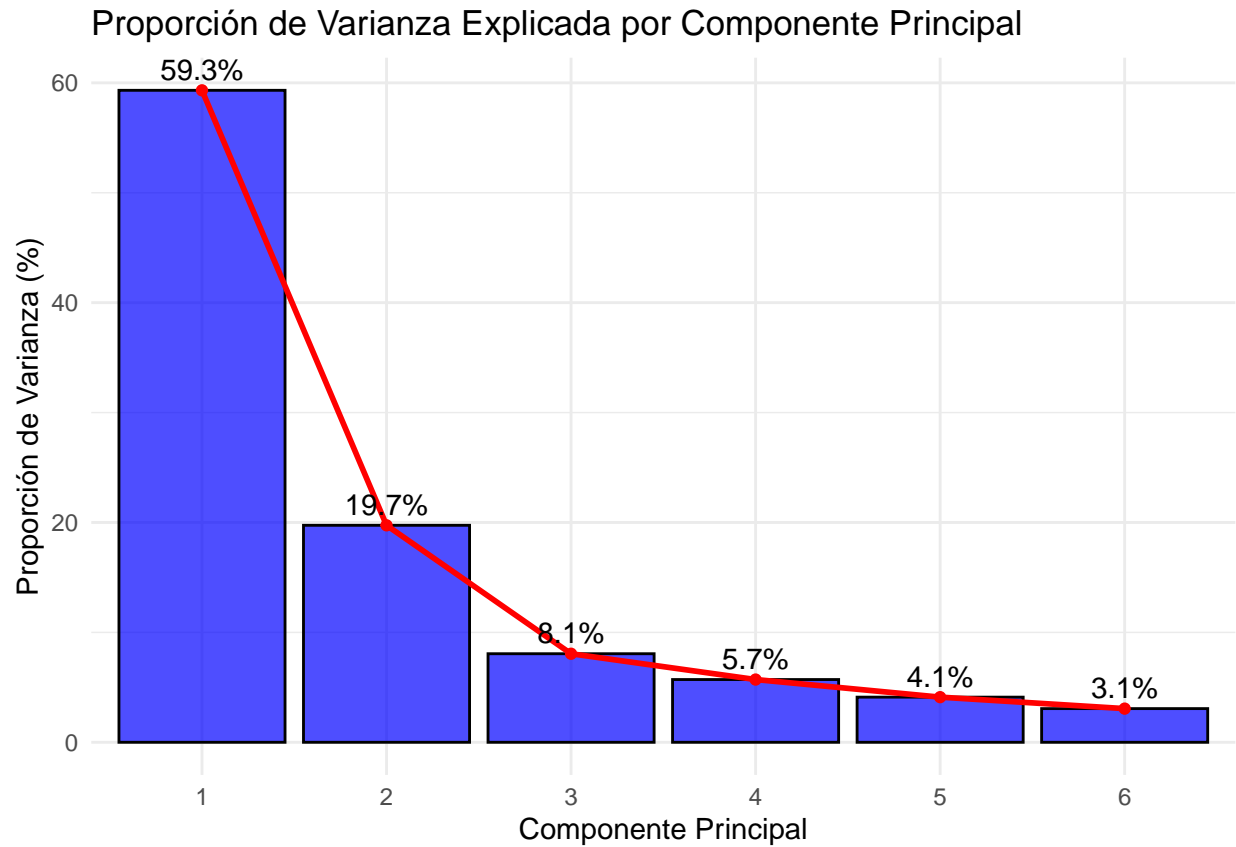
```
## 'geom_line()': Each group consists of only one observation.  
## i Do you need to adjust the group aesthetic?
```





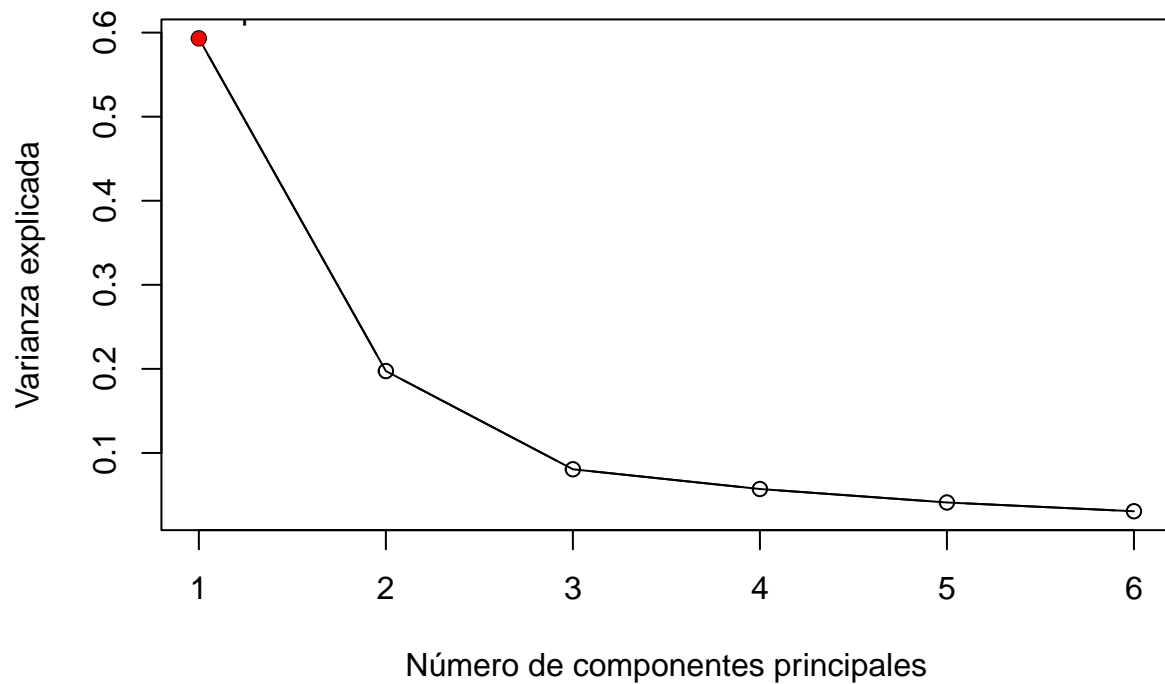
```
## Warning in geom_point(color = "red", linewidth = 2): Ignoring unknown
## parameters: 'linewidth'
```

```
## Warning in geom_text(aes(label = paste0(round(Proportion_of_Variance, 1), :
## Ignoring unknown parameters: 'linewidth'
```



Ahora bien, basándonos en lo visto en el curso, es posible aplicar la “*regla del codo*” en el análisis de componentes principales. En esta etapa se calculará la varianza explicada por cada componente principal y se trazaría un gráfico para identificar el punto en el que se produce un cambio significativo en la pendiente (el “*codo*”), en este se observa que la curva se genera para el primer y segundo componente para confirmar lo observado anteriormente.

Regla del codo para el PCA

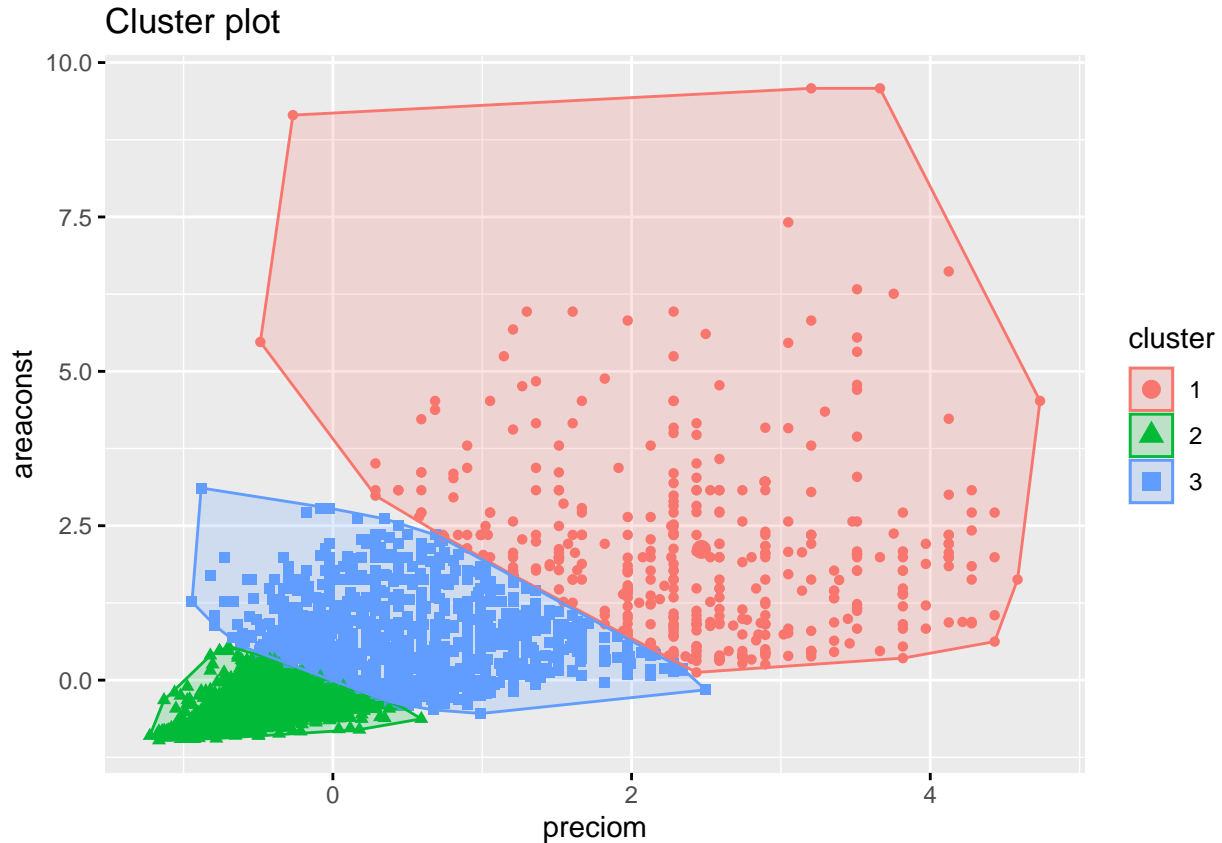


2. Análisis de Conglomerados (Clustering):

```
## Warning: package 'factoextra' was built under R version 4.4.2
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
##      preciom  areaconst
## 1  2.4677661  2.1167169
## 2 -0.5434125 -0.5423895
## 3  0.4802026  0.5818992
```

Los resultados del análisis previo, muestran tres centroides correspondientes a los clusters identificados por el algoritmo *K-means*, utilizando las variables **precio** (precio por metro cuadrado) y **areaconst** (área construida). Los centroides de los clusters pueden interpretarse de la siguiente manera:

- **Cluster 1:** Se caracteriza por valores moderadamente bajos tanto en **precio** como en **areaconst**.
- **Cluster 2:** Se distingue por valores moderadamente altos en ambas variables, **precio** y **areaconst**.
- **Cluster 3:** Está caracterizado por valores relativamente altos en **precio** y **areaconst**.

Este análisis permite observar cómo se agrupan las viviendas según estas dos variables clave, proporcionando una comprensión más clara de las distintas categorías en el mercado inmobiliario.

3. Análisis de Correspondencia:

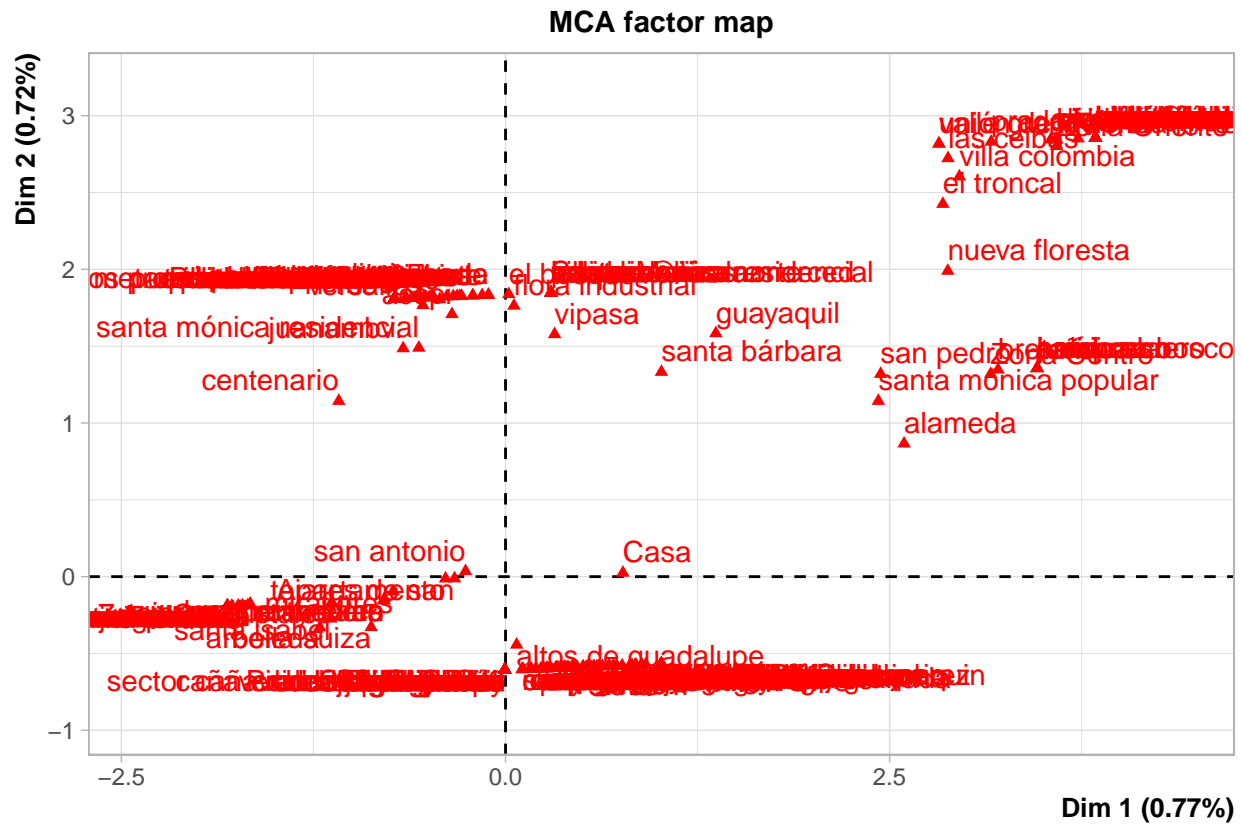
El análisis de correspondencia múltiple (ACM) realizado proporciona valiosa información sobre las interacciones entre las variables categóricas en el conjunto de datos. A continuación se presenta un resumen de las dimensiones principales identificadas, que reflejan la estructura subyacente de los datos y la cantidad de varianza total que explican:

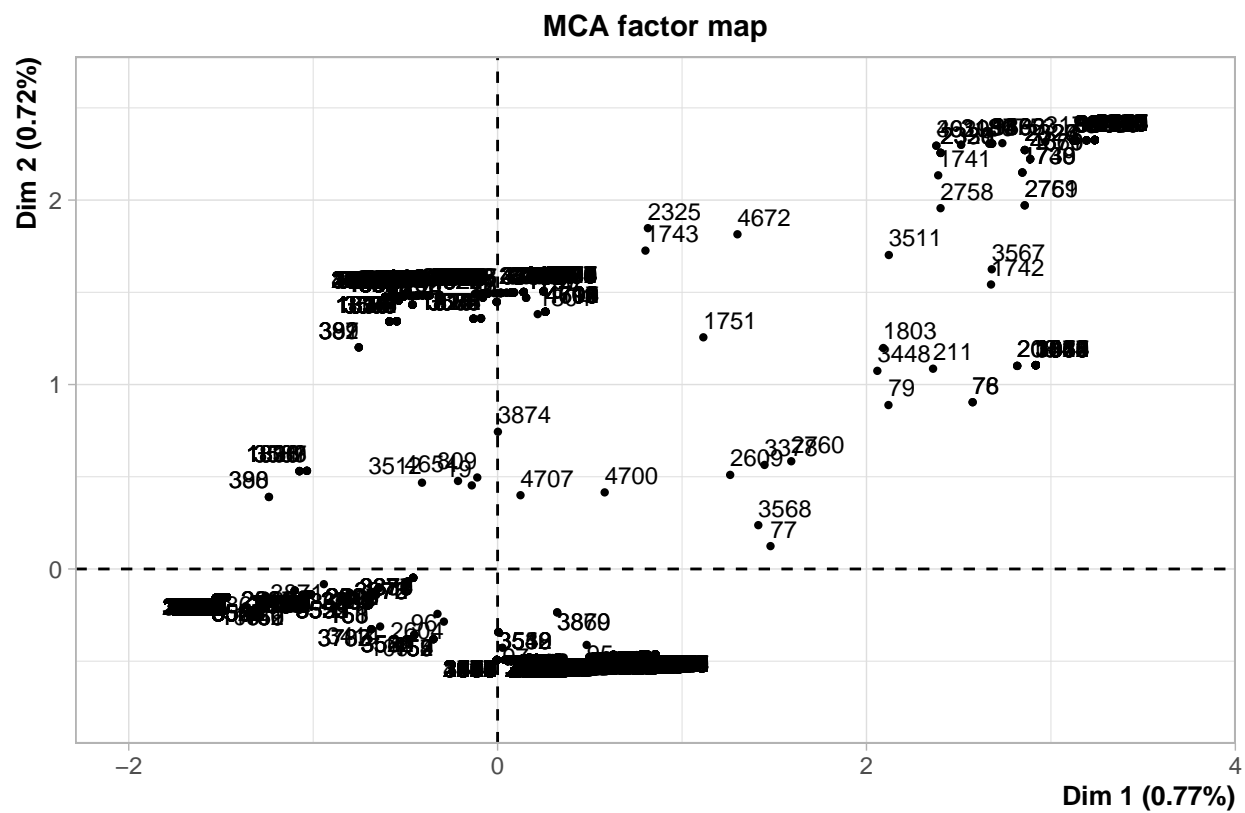
- **Dimensión 1:** Con una varianza explicada de 0.711, representando el 77.0% de la varianza total, esta dimensión muestra una relación entre el tipo de propiedad (apartamento o casa) y la zona geográfica (centro, norte, oeste, oriente, sur), con contrastes significativos en los valores.
- **Dimensión 2:** Explicando el 66.3% de la varianza (0.718% de la varianza total), esta dimensión destaca la relación entre la zona geográfica y otras variables categóricas como “3 de julio”, “acopi”, y “aguablanca”, mostrando asociaciones adicionales.

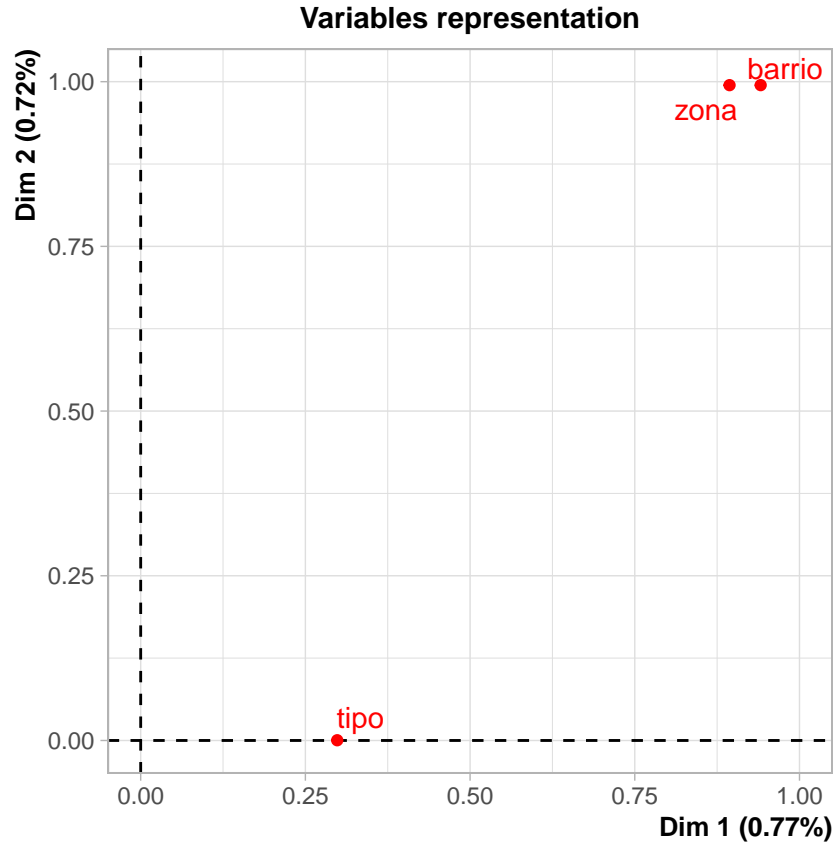
- **Dimensión 3:** Con una varianza explicada de 0.658 (0.713% de la varianza total), esta dimensión complementa la información sobre la distribución espacial y las características de ciertas zonas en relación con otras variables categóricas.

El análisis también revela cómo cada categoría contribuye a las dimensiones principales y su asociación con otras categorías. Por ejemplo, la categoría “Casa” se relaciona positivamente con la **Dimensión 1**, mientras que “Zona Sur” presenta una fuerte correlación positiva con la **Dimensión 2**, reflejando las tendencias de agrupamiento geográfico y las características del tipo de propiedad en las diferentes zonas.

Warning: package 'FactoMineR' was built under R version 4.4.2







```
##
## Call:
## MCA(X = df_corr)
##
## Eigenvalues
##
```

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7
## Variance	0.711	0.663	0.658	0.645	0.446	0.333	0.333
## % of var.	0.770	0.718	0.713	0.699	0.483	0.361	0.361
## Cumulative % of var.	0.770	1.488	2.201	2.900	3.383	3.744	4.105

```
##
```

	Dim.8	Dim.9	Dim.10	Dim.11	Dim.12	Dim.13	Dim.14
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	4.466	4.827	5.188	5.549	5.910	6.271	6.633

```
##
```

	Dim.15	Dim.16	Dim.17	Dim.18	Dim.19	Dim.20	Dim.21
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	6.994	7.355	7.716	8.077	8.438	8.799	9.160

```
##
```

	Dim.22	Dim.23	Dim.24	Dim.25	Dim.26	Dim.27	Dim.28
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	9.521	9.882	10.243	10.604	10.965	11.326	11.687

```
##
```

	Dim.29	Dim.30	Dim.31	Dim.32	Dim.33	Dim.34	Dim.35
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	12.048	12.409	12.770	13.131	13.492	13.853	14.214

##	Dim.36	Dim.37	Dim.38	Dim.39	Dim.40	Dim.41	Dim.42
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	14.575	14.936	15.297	15.658	16.019	16.380	16.741
##	Dim.43	Dim.44	Dim.45	Dim.46	Dim.47	Dim.48	Dim.49
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	17.102	17.463	17.824	18.185	18.546	18.907	19.268
##	Dim.50	Dim.51	Dim.52	Dim.53	Dim.54	Dim.55	Dim.56
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	19.629	19.990	20.351	20.712	21.073	21.434	21.795
##	Dim.57	Dim.58	Dim.59	Dim.60	Dim.61	Dim.62	Dim.63
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	22.156	22.517	22.878	23.239	23.600	23.961	24.322
##	Dim.64	Dim.65	Dim.66	Dim.67	Dim.68	Dim.69	Dim.70
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	24.683	25.044	25.405	25.766	26.127	26.488	26.849
##	Dim.71	Dim.72	Dim.73	Dim.74	Dim.75	Dim.76	Dim.77
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	27.210	27.571	27.932	28.293	28.654	29.015	29.376
##	Dim.78	Dim.79	Dim.80	Dim.81	Dim.82	Dim.83	Dim.84
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	29.737	30.098	30.459	30.820	31.181	31.542	31.903
##	Dim.85	Dim.86	Dim.87	Dim.88	Dim.89	Dim.90	Dim.91
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	32.264	32.625	32.986	33.347	33.708	34.069	34.430
##	Dim.92	Dim.93	Dim.94	Dim.95	Dim.96	Dim.97	Dim.98
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	34.791	35.152	35.513	35.874	36.235	36.596	36.957
##	Dim.99	Dim.100	Dim.101	Dim.102	Dim.103	Dim.104	Dim.105
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	37.318	37.679	38.040	38.401	38.762	39.123	39.484
##	Dim.106	Dim.107	Dim.108	Dim.109	Dim.110	Dim.111	Dim.112
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	39.846	40.207	40.568	40.929	41.290	41.651	42.012
##	Dim.113	Dim.114	Dim.115	Dim.116	Dim.117	Dim.118	Dim.119
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	42.373	42.734	43.095	43.456	43.817	44.178	44.539
##	Dim.120	Dim.121	Dim.122	Dim.123	Dim.124	Dim.125	Dim.126
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	44.900	45.261	45.622	45.983	46.344	46.705	47.066
##	Dim.127	Dim.128	Dim.129	Dim.130	Dim.131	Dim.132	Dim.133
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333

## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	47.427	47.788	48.149	48.510	48.871	49.232	49.593
##	Dim.134	Dim.135	Dim.136	Dim.137	Dim.138	Dim.139	Dim.140
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	49.954	50.315	50.676	51.037	51.398	51.759	52.120
##	Dim.141	Dim.142	Dim.143	Dim.144	Dim.145	Dim.146	Dim.147
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	52.481	52.842	53.203	53.564	53.925	54.286	54.647
##	Dim.148	Dim.149	Dim.150	Dim.151	Dim.152	Dim.153	Dim.154
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	55.008	55.369	55.730	56.091	56.452	56.813	57.174
##	Dim.155	Dim.156	Dim.157	Dim.158	Dim.159	Dim.160	Dim.161
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	57.535	57.896	58.257	58.618	58.979	59.340	59.701
##	Dim.162	Dim.163	Dim.164	Dim.165	Dim.166	Dim.167	Dim.168
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	60.062	60.423	60.784	61.145	61.506	61.867	62.228
##	Dim.169	Dim.170	Dim.171	Dim.172	Dim.173	Dim.174	Dim.175
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	62.589	62.950	63.311	63.672	64.033	64.394	64.755
##	Dim.176	Dim.177	Dim.178	Dim.179	Dim.180	Dim.181	Dim.182
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	65.116	65.477	65.838	66.199	66.560	66.921	67.282
##	Dim.183	Dim.184	Dim.185	Dim.186	Dim.187	Dim.188	Dim.189
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	67.643	68.004	68.365	68.726	69.087	69.448	69.809
##	Dim.190	Dim.191	Dim.192	Dim.193	Dim.194	Dim.195	Dim.196
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	70.170	70.531	70.892	71.253	71.614	71.975	72.336
##	Dim.197	Dim.198	Dim.199	Dim.200	Dim.201	Dim.202	Dim.203
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	72.697	73.059	73.420	73.781	74.142	74.503	74.864
##	Dim.204	Dim.205	Dim.206	Dim.207	Dim.208	Dim.209	Dim.210
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	75.225	75.586	75.947	76.308	76.669	77.030	77.391
##	Dim.211	Dim.212	Dim.213	Dim.214	Dim.215	Dim.216	Dim.217
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	77.752	78.113	78.474	78.835	79.196	79.557	79.918
##	Dim.218	Dim.219	Dim.220	Dim.221	Dim.222	Dim.223	Dim.224
## Variance	0.333	0.333	0.333	0.333	0.333	0.333	0.333
## % of var.	0.361	0.361	0.361	0.361	0.361	0.361	0.361
## Cumulative % of var.	80.279	80.640	81.001	81.362	81.723	82.084	82.445

```

##          Dim.225 Dim.226 Dim.227 Dim.228 Dim.229 Dim.230 Dim.231
## Variance          0.333   0.333   0.333   0.333   0.333   0.333   0.333
## % of var.          0.361   0.361   0.361   0.361   0.361   0.361   0.361
## Cumulative % of var. 82.806  83.167  83.528  83.889  84.250  84.611  84.972
##          Dim.232 Dim.233 Dim.234 Dim.235 Dim.236 Dim.237 Dim.238
## Variance          0.333   0.333   0.333   0.333   0.333   0.333   0.333
## % of var.          0.361   0.361   0.361   0.361   0.361   0.361   0.361
## Cumulative % of var. 85.333  85.694  86.055  86.416  86.777  87.138  87.499
##          Dim.239 Dim.240 Dim.241 Dim.242 Dim.243 Dim.244 Dim.245
## Variance          0.333   0.333   0.333   0.333   0.333   0.333   0.333
## % of var.          0.361   0.361   0.361   0.361   0.361   0.361   0.361
## Cumulative % of var. 87.860  88.221  88.582  88.943  89.304  89.665  90.026
##          Dim.246 Dim.247 Dim.248 Dim.249 Dim.250 Dim.251 Dim.252
## Variance          0.333   0.333   0.333   0.333   0.333   0.333   0.333
## % of var.          0.361   0.361   0.361   0.361   0.361   0.361   0.361
## Cumulative % of var. 90.387  90.748  91.109  91.470  91.831  92.192  92.553
##          Dim.253 Dim.254 Dim.255 Dim.256 Dim.257 Dim.258 Dim.259
## Variance          0.333   0.333   0.333   0.333   0.333   0.333   0.333
## % of var.          0.361   0.361   0.361   0.361   0.361   0.361   0.361
## Cumulative % of var. 92.914  93.275  93.636  93.997  94.358  94.719  95.080
##          Dim.260 Dim.261 Dim.262 Dim.263 Dim.264 Dim.265 Dim.266
## Variance          0.333   0.333   0.333   0.333   0.333   0.333   0.333
## % of var.          0.361   0.361   0.361   0.361   0.361   0.361   0.361
## Cumulative % of var. 95.441  95.802  96.163  96.524  96.885  97.246  97.607
##          Dim.267 Dim.268 Dim.269 Dim.270 Dim.271 Dim.272 Dim.273
## Variance          0.333   0.333   0.333   0.333   0.333   0.333   0.167
## % of var.          0.361   0.361   0.361   0.361   0.361   0.361   0.180
## Cumulative % of var. 97.968  98.329  98.690  99.051  99.412  99.773  99.954
##          Dim.274 Dim.275 Dim.276 Dim.277
## Variance          0.023   0.008   0.008   0.003
## % of var.          0.025   0.009   0.009   0.003
## Cumulative % of var. 99.979  99.988  99.997 100.000
##
## Individuals (the 10 first)
##          Dim.1   ctr   cos2   Dim.2   ctr   cos2   Dim.3   ctr
## 1          | 0.855 0.021 0.000 | -0.463 0.007 0.000 | -0.200 0.001
## 2          | -0.461 0.006 0.003 | 1.433 0.064 0.026 | -0.794 0.020
## 3          | -0.461 0.006 0.003 | 1.433 0.064 0.026 | -0.794 0.020
## 4          | -0.461 0.006 0.003 | 1.433 0.064 0.026 | -0.794 0.020
## 5          | -0.461 0.006 0.003 | 1.433 0.064 0.026 | -0.794 0.020
## 6          | -0.005 0.000 0.000 | 1.448 0.066 0.027 | -0.792 0.020
## 7          | -0.005 0.000 0.000 | 1.448 0.066 0.027 | -0.792 0.020
## 8          | -0.005 0.000 0.000 | 1.448 0.066 0.027 | -0.792 0.020
## 9          | -0.005 0.000 0.000 | 1.448 0.066 0.027 | -0.792 0.020
## 10         | -0.461 0.006 0.003 | 1.433 0.064 0.026 | -0.794 0.020
##          cos2
## 1          0.000 |
## 2          0.008 |
## 3          0.008 |
## 4          0.008 |
## 5          0.008 |
## 6          0.008 |
## 7          0.008 |
## 8          0.008 |

```

```

## 9          0.008 |
## 10         0.008 |
##
## Categories (the 10 first)
##          Dim.1    ctr    cos2  v.test    Dim.2    ctr    cos2
## Apartamento | -0.390  4.729  0.298 -37.867 | -0.013  0.005  0.000
## Casa        |  0.764  9.254  0.298  37.867 |  0.025  0.010  0.000
## Zona Centro |  3.157  3.499  0.075  19.013 |  1.318  0.654  0.013
## Zona Norte  | -0.428  1.584  0.041 -14.112 |  1.805 30.162  0.736
## Zona Oeste  | -1.661 20.262  0.512 -49.634 | -0.177  0.246  0.006
## Zona Oriente |  3.587 12.169  0.265  35.687 |  2.802  7.959  0.162
## Zona Sur    |  0.385  4.387  0.254  34.936 | -0.588 10.971  0.592
## 3 de julio   |  1.014  0.010  0.000  1.014 | -0.569  0.003  0.000
## acopi        | -0.349  0.025  0.001 -1.601 |  1.708  0.640  0.013
## aguablanca   | -0.005  0.000  0.000 -0.005 | -0.607  0.004  0.000
##          v.test    Dim.3    ctr    cos2  v.test
## Apartamento -1.230 | -0.002  0.000  0.000 -0.172 |
## Casa        1.230 |  0.003  0.000  0.000  0.172 |
## Zona Centro 7.938 |  1.385  0.727  0.014  8.340 |
## Zona Norte  59.470 | -0.970  8.784  0.213 -31.973 |
## Zona Oeste  -5.281 |  1.470 17.137  0.401 43.921 |
## Zona Oriente 27.874 |  4.582 21.443  0.432 45.581 |
## Zona Sur    -53.359 | -0.244  1.908  0.102 -22.167 |
## 3 de julio   -0.569 | -0.247  0.001  0.000 -0.247 |
## acopi        7.843 | -0.960  0.204  0.004 -4.408 |
## aguablanca   -0.607 | -0.252  0.001  0.000 -0.252 |
##
## Categorical variables (eta2)
##          Dim.1 Dim.2 Dim.3
## tipo      | 0.298 0.000 0.000 |
## zona      | 0.894 0.995 0.987 |
## barrio    | 0.941 0.995 0.987 |

```

Conclusiones

A partir del análisis realizado, se pueden extraer las siguientes conclusiones clave:

1. **Asociación entre tamaño, ubicación y precio:** Se observó una fuerte relación entre el tamaño de la vivienda, su ubicación geográfica y el precio. Esto sugiere que tanto el área construida como la zona donde se encuentra la propiedad son factores determinantes en la fijación de precios en el mercado inmobiliario.
2. **Segmentación de mercado:** El análisis de conglomerados permitió identificar distintos segmentos de mercado, lo que abre la puerta a la creación de estrategias de marketing personalizadas. Al comprender las características de cada grupo de viviendas, es posible ofrecer soluciones más específicas y dirigidas a las preferencias de cada segmento.
3. **Variación de preferencias según zona geográfica:** Las preferencias del mercado varían considerablemente según la zona geográfica y el barrio. Esto resalta la importancia de adaptar la oferta de viviendas a las necesidades y demandas específicas de cada área, lo que podría mejorar la competitividad y la satisfacción del cliente en el sector inmobiliario.

Repositorio con código