

# EDA\_MICali

Jhon Alexander Rojas Tavera

2023-08-13

## Contents

1. Introducción . . . . .	1
2. Objetivos . . . . .	2
3. Métodos . . . . .	2
4. Resultados . . . . .	9
5. Discusión . . . . .	9
6. Conclusiones . . . . .	9
7. Anexos . . . . .	9

```
##
## Attaching package: 'bs4Dash'

## The following object is masked from 'package:graphics':
##
##     box
```

Informe Unidad I, Analisis exploratorio de datos

## 1. Introducción

El análisis del mercado inmobiliario es una herramienta crucial para comprender la dinámica y las tendencias que impactan en el sector de bienes raíces. En este informe, se abordará el análisis del mercado inmobiliario en Cali, Colombia, utilizando datos recopilados por la empresa B&C (Bienes y Casas), una agencia de bienes raíces de renombre en la ciudad.

Cali, una ciudad en constante evolución, ha sido testigo de un crecimiento sustancial en el mercado inmobiliario en los últimos años. Este crecimiento está impulsado por diversos factores, como el aumento de la población, la inversión extranjera directa y la construcción de nuevos proyectos residenciales. Con el propósito de aprovechar este entorno dinámico, la empresa B&C ha recopilado y sistematizado una vasta cantidad de datos sobre viviendas en Cali, que incluyen información sobre precios, ubicaciones, características y detalles de ventas.

Este informe tiene como objetivo explorar y analizar estos datos con el fin de identificar tendencias y patrones significativos que puedan servir como base para la toma de decisiones estratégicas por parte de la empresa B&C. A través de un enfoque descriptivo y analítico, se espera brindar una visión completa y precisa del mercado inmobiliario en Cali, permitiendo a la empresa tomar medidas informadas para su crecimiento y éxito continuo.

El informe está estructurado en varios apartados, que incluyen objetivos, métodos utilizados, resultados obtenidos, discusión de los hallazgos, conclusiones y anexos. Cada sección proporcionará una perspectiva detallada y significativa para comprender mejor la dinámica del mercado inmobiliario en Cali.

## 2. Objetivos

Objetivo general:

Texto 1

Objetivo específico 1:

Texto 1

Específico 2:

Texto 1

## 3. Métodos

Lectura de la base de datos desde archivo CSV, en mi caso la hice mediante el uso del fichero remitido por la profesora denominado “vivienda\_faltantes.csv”, con las siguientes acciones en RStudio: File, Import DataSet, From Text(readr) y el nombre del archivo se paso como atributo a la variable “Baseviviendas”:

```
Baseviviendas <- read_csv("vivienda_faltantes.csv")
```

```
## Rows: 8330 Columns: 13
## -- Column specification -----
## Delimiter: ","
## chr (3): zona, tipo, barrio
## dbl (10): id, piso, estrato, preciom, areaconst, parquea, banios, habitac, l...
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
head(Baseviviendas)
```

```
## # A tibble: 6 x 13
##       id zona      piso estrato preciom areaconst parquea banios habitac tipo
##   <dbl> <chr>    <dbl>   <dbl>   <dbl>    <dbl>   <dbl>  <dbl>  <chr>
## 1  8312 Zona Oeste      4       6   1300     318      2      4      2 Apart~
## 2  8311 Zona Oeste      1       6    480     300      1      4      4 Casa
## 3  8307 Zona Oeste     NA       5   1200     800      4      7      5 Casa
## 4  8296 Zona Sur       2       3    220     150      1      2      4 Casa
## 5  8297 Zona Oeste     NA       5    330     112      2      4      3 Casa
## 6  8298 Zona Sur      NA       5   1350     390      8     10     10 Casa
## # i 3 more variables: barrio <chr>, longitud <dbl>, latitud <dbl>
```

```
Baseviviendas_resumen <- summary(Baseviviendas)
Baseviviendas_resumen
```

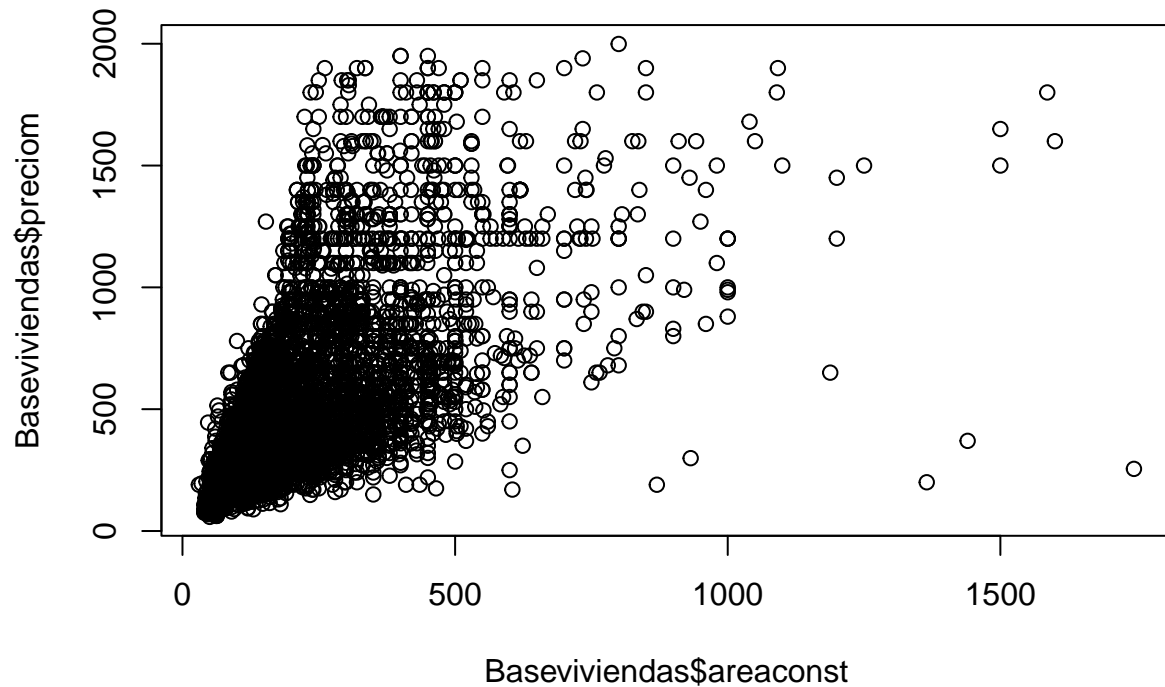
```
##      id      zona      piso      estrato
## Min.   : 1   Length:8330   Min.   : 1.000   Min.   :3.000
## 1st Qu.:2082   Class :character 1st Qu.: 2.000   1st Qu.:4.000
## Median :4164   Mode  :character  Median : 3.000   Median :5.000
## Mean   :4164                      Mean  : 3.772   Mean   :4.634
## 3rd Qu.:6246                      3rd Qu.: 5.000   3rd Qu.:5.000
## Max.   :8319                      Max.   :12.000   Max.   :6.000
## NA's   :3                        NA's   :2641    NA's   :3
##      preciom      areaconst      parquea      banios
## Min.   : 58.0   Min.   : 30   Min.   : 1.000   Min.   : 0.000
## 1st Qu.: 220.0   1st Qu.: 80   1st Qu.: 1.000   1st Qu.: 2.000
## Median : 330.0   Median : 123   Median : 2.000   Median : 3.000
## Mean   : 434.2   Mean   : 175   Mean   : 1.836   Mean   : 3.112
## 3rd Qu.: 540.0   3rd Qu.: 229   3rd Qu.: 2.000   3rd Qu.: 4.000
## Max.   :1999.0   Max.   :1745   Max.   :10.000   Max.   :10.000
## NA's   :2       NA's   :3       NA's   :1606    NA's   :3
##      habitac      tipo      barrio      longitud
## Min.   : 0.000   Length:8330   Length:8330   Min.   : -76576.00
## 1st Qu.: 3.000   Class :character  Class :character 1st Qu.: -76506.00
## Median : 3.000   Mode  :character  Mode  :character  Median :   -76.54
## Mean   : 3.605                      Mean   : -21845.13
## 3rd Qu.: 4.000                      3rd Qu.:   -76.52
## Max.   :10.000                      Max.   :   -76.46
## NA's   :3                        NA's   :3
##      latitud
## Min.   : 3.333
## 1st Qu.: 3.390
## Median : 3.450
## Mean   : 970.370
## 3rd Qu.:3367.000
## Max.   :3497.000
## NA's   :3
```

```
# Filtrar los datos para mostrar solo las primeras 10 filas y las columnas 'precio' y 'zona'
Baseviviendas_subset <- Baseviviendas[1:10, c('preciom', 'zona')]

# Mostrar el conjunto de datos filtrado
Baseviviendas_subset
```

```
## # A tibble: 10 x 2
##   preciom zona
##   <dbl> <chr>
## 1   1300 Zona Oeste
## 2    480 Zona Oeste
## 3   1200 Zona Oeste
## 4    220 Zona Sur
## 5    330 Zona Oeste
## 6   1350 Zona Sur
## 7    305 Zona Sur
## 8    480 Zona Oeste
## 9    275 Zona Sur
## 10   285 Zona Sur
```

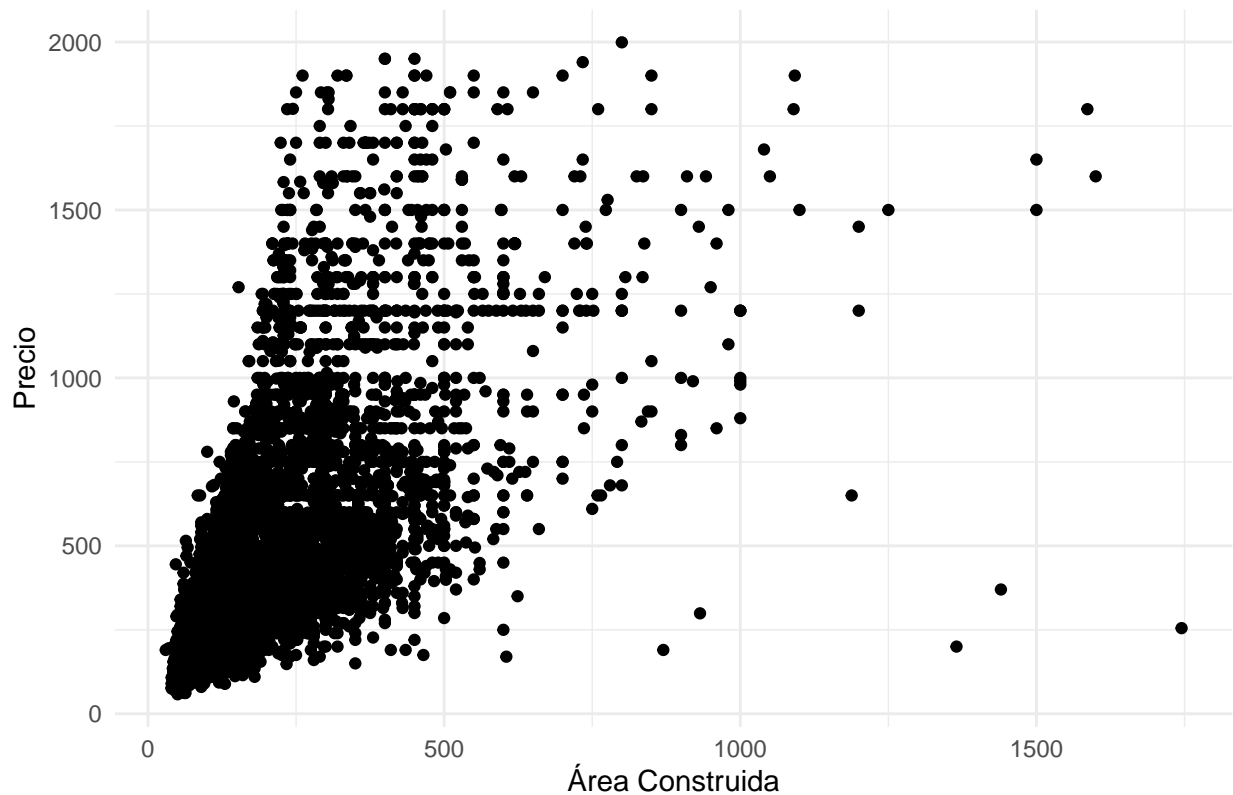
```
# Crear un gráfico de dispersión entre 'areaconst' y 'preciom'  
plot(Baseviviendas$areaconst, Baseviviendas$preciom)
```



```
ggplot(Baseviviendas, aes(x = areaconst, y = preciom)) +  
  geom_point() +  
  labs(title = "Relación entre Área Construida y Precio",  
        x = "Área Construida",  
        y = "Precio") +  
  theme_minimal()
```

```
## Warning: Removed 3 rows containing missing values ('geom_point()').
```

## Relación entre Área Construida y Precio



```
# Calcular la cantidad de valores faltantes en cada columna y asignarlos a EDA_MICali_faltantes
Baseviviendas_faltantes <- colSums(is.na(Baseviviendas))
```

```
# Mostrar la cantidad de valores faltantes
Baseviviendas_faltantes
```

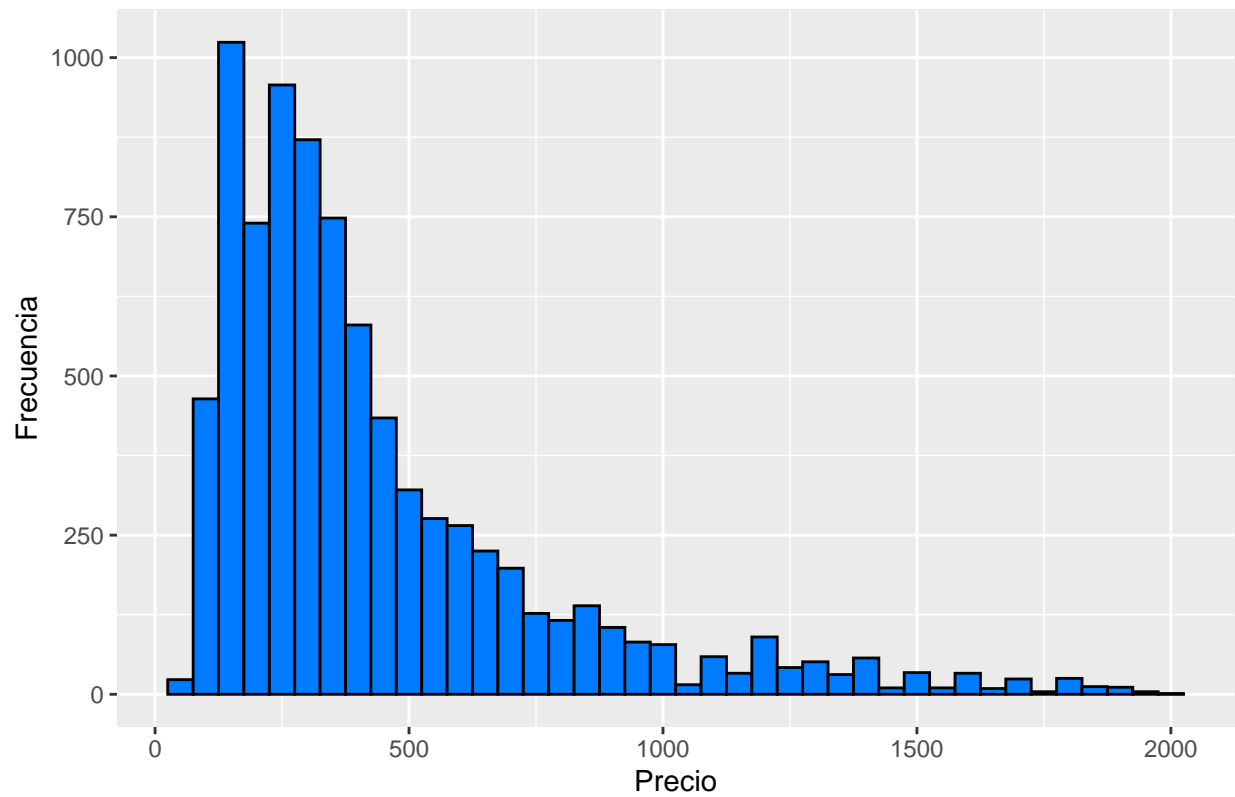
```
##      id      zona      piso  estrato  preciom areaconst  parquea  banios
##      3         3      2641        3        2          3      1606        3
## habitac      tipo  barrio longitud  latitud
##      3         3        3         3         3
```

```
# Histograma de Precios
```

```
ggplot(Baseviviendas, aes(x = preciom)) +
  geom_histogram(binwidth = 50, fill = "#007bff", color = "black") +
  labs(title = "Distribución de Precios",
       x = "Precio",
       y = "Frecuencia")
```

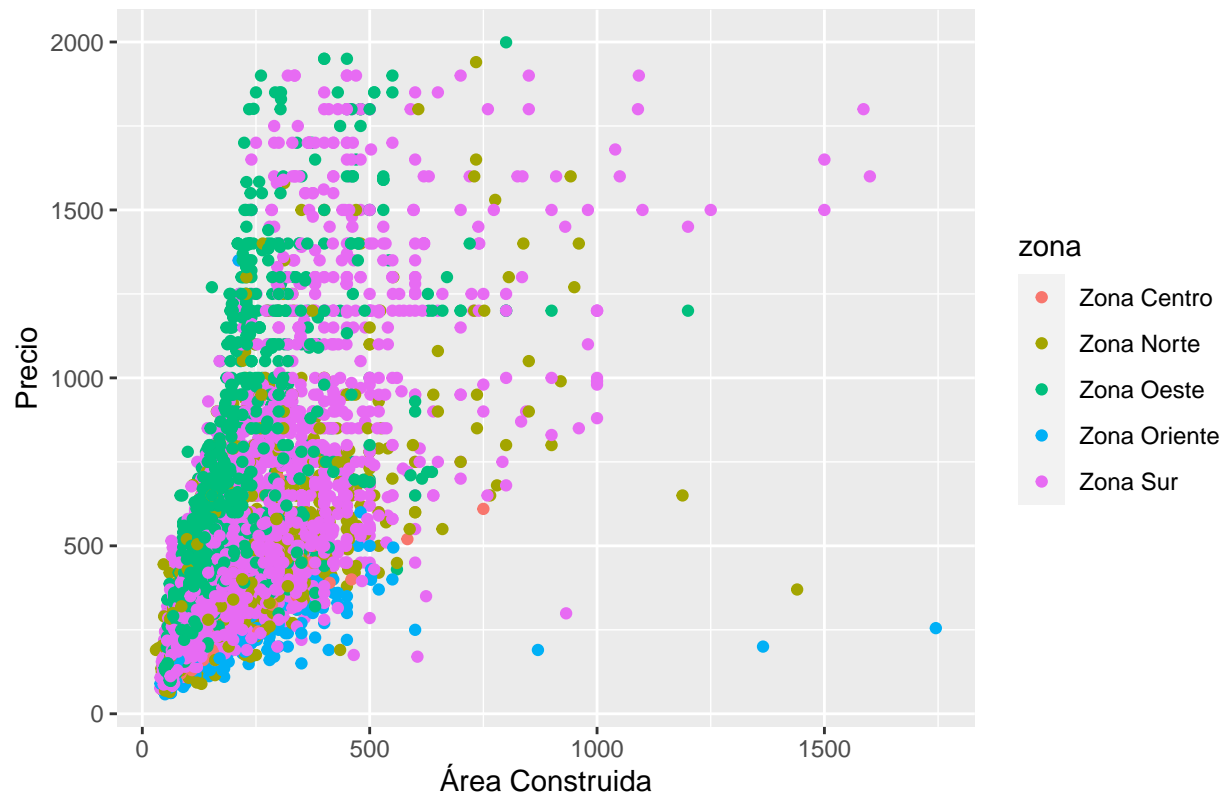
```
## Warning: Removed 2 rows containing non-finite values ('stat_bin()').
```

Distribución de Precios



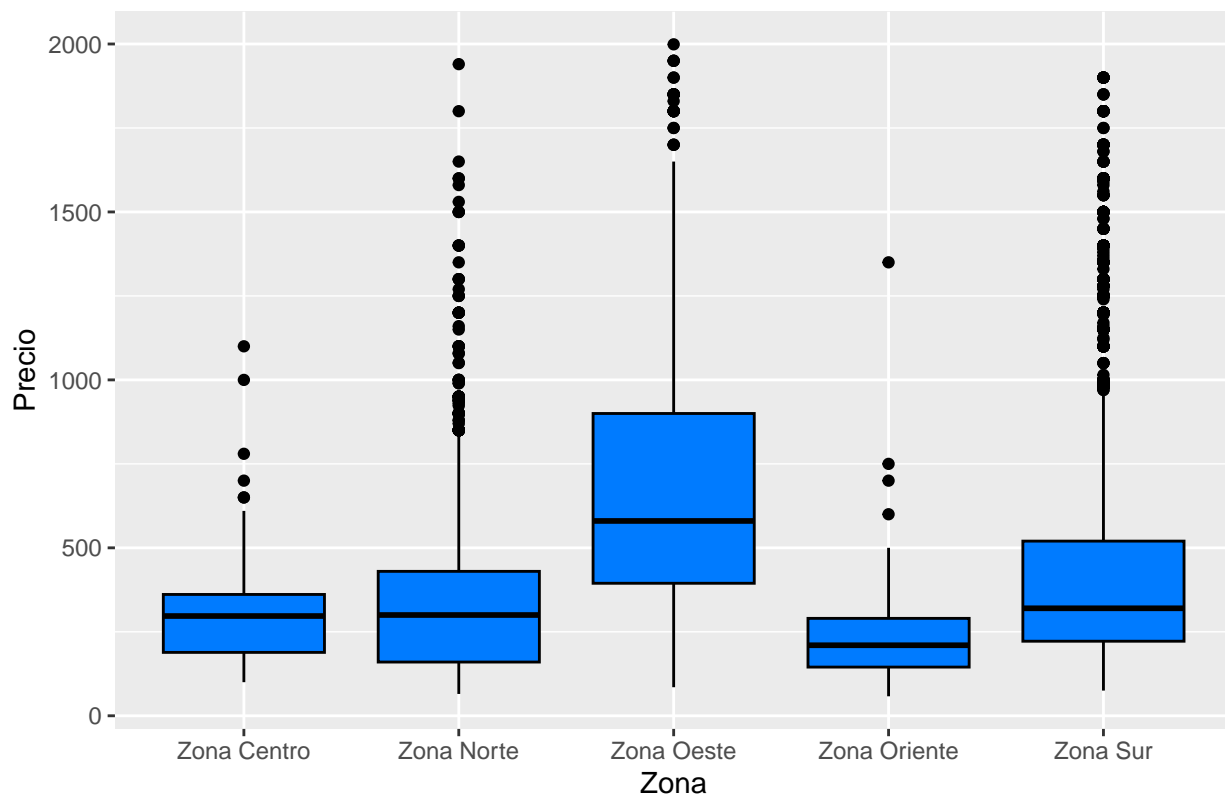
```
# Gráfico de Dispersión con Color por Zona (sin valores NA)
ggplot(Baseviviendas %>% drop_na(zona), aes(x = areaconst, y = preciom, color = zona)) +
  geom_point() +
  labs(title = "Relación entre Área Construida y Precio por Zona",
       x = "Área Construida",
       y = "Precio")
```

Relación entre Área Construida y Precio por Zona



```
# Boxplot de Precios por Zona (sin valores NA)
ggplot(Baseviviendas %>% drop_na(zona), aes(x = zona, y = preciom)) +
  geom_boxplot(fill = "#007bff", color = "black") +
  labs(title = "Distribución de Precios por Zona",
       x = "Zona",
       y = "Precio")
```

Distribución de Precios por Zona



```
# Instalar el paquete si no está instalado
if (!requireNamespace("corrplot", quietly = TRUE)) {
  install.packages("corrplot")
}
```

```
# Cargar el paquete
library(corrplot)
```

```
## corrplot 0.92 loaded
```

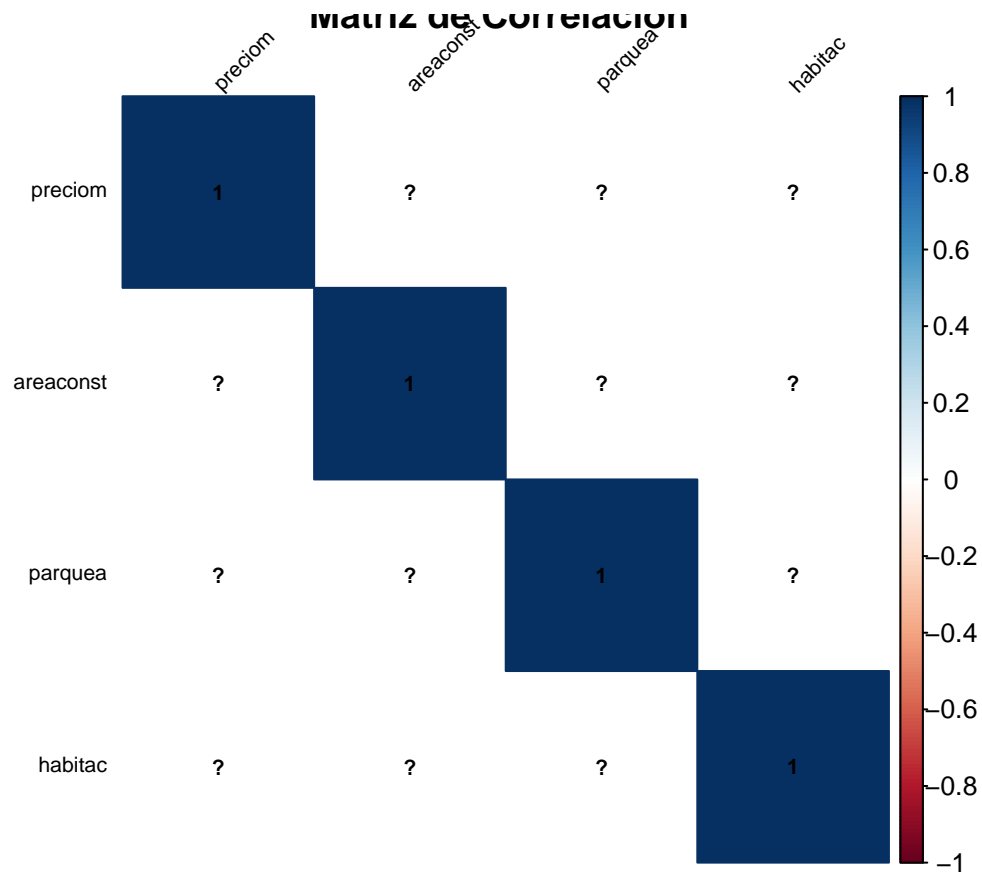
```
# Cálculo de la Matriz de Correlación
```

```
correlation_matrix <- cor(Baseviviendas[, c("preciom", "areaconst", "parquea", "habitac")])
```

```
# Visualización de la Matriz de Correlación usando corrplot()
```

```
corrplot(correlation_matrix,
  method = "color", # Puedes ajustar el método de visualización
  tl.col = "black",
  tl.srt = 45,
  addCoef.col = "black", # Color de los coeficientes de correlación
  number.cex = 0.7, # Tamaño de los números en la visualización
  tl.cex = 0.7, # Tamaño de las etiquetas en los ejes
  title = "Matriz de Correlación")
```





4. Resultados

5. Discusión

6. Conclusiones

7. Anexos