

# Actividad\_03

Juan José Restrepo Rosero

23 September 2024

## Contents

Cargar las librerías necesarias para el análisis exploratorio 1

Cargar los datos del dataframe “vivienda4” si aún no se han cargado 1

Definición del problema

Con base en los datos de ofertas de vivienda descargadas del portal Fincaraiz para apartamento de estrato 4 con área construida menor a  $200 \text{ m}^2$  (vivienda4.RDS) la inmobiliaria A&C requiere el apoyo de un científico de datos en la construcción de un modelo que lo oriente sobre los precios de inmuebles. Con este propósito el equipo de asesores a diseñado los siguientes pasos para obtener un modelo y así poder a futuro determinar los precios de los inmuebles a negociar

**1. Realice un análisis exploratorio de las variables precio de vivienda (millones de pesos COP) y área de la vivienda (metros cuadrados) - incluir gráficos e indicadores apropiados interpretados.**

```
data(vivienda4)
```

## Cargar las librerías necesarias para el análisis exploratorio

```
library(ggplot2) # Para graficar library(dplyr) # Para manipulación de datos library(summarytools) # Para resúmenes descriptivos
```

## Cargar los datos del dataframe “vivienda4” si aún no se han cargado

```
if (!exists("vivienda4")) { data(vivienda4) }
```

```
# 1. Validar datos faltantes por variable
missing_data <- sapply(vivienda4, function(x) sum(is.na(x)))
cat("Datos faltantes por variable:\n")
```

```
## Datos faltantes por variable:
```

```
print(missing_data)
```

```
##      zona   estrato   preciom areaconst   tipo
##      0         0         0         0         0
```

```
# 2. Validar si existen datos vacíos o null dentro de las variables del dataframe
if (any(is.na(vivienda4))) {
  cat("El dataframe contiene valores nulos o vacíos.\n")
}
```

```
# 3. Cuantificar valores duplicados y eliminarlos
duplicated_count <- sum(duplicated(vivienda4))
cat("Número de filas duplicadas:", duplicated_count, "\n")
```

```
## Número de filas duplicadas: 0
```

```
# Eliminar filas duplicadas
vivienda4 <- unique(vivienda4)
```

```
# 4. Dejar el dataframe final sin outliers (valores atípicos)
# Puedes utilizar un criterio específico para definir outliers, por ejemplo, basado en desviaciones est
# Para este ejemplo, se eliminarán observaciones con valores de "preciom" que estén a más de 3 desviaci

# Calcular la media y desviación estándar de la variable "preciom"
media_precio <- mean(vivienda4$preciom)
desviacion_precio <- sd(vivienda4$preciom)
```

```
# Crear un vector lógico unidimensional para filtrar los outliers
condicion_outliers <- abs((vivienda4$preciom - media_precio) / desviacion_precio) <= 3
```

```
vivienda4 <- vivienda4 %>%
  filter(condicion_outliers)
print(vivienda4)
```

```
## # A tibble: 1,687 x 5
##   zona      estrato preciom areaconst tipo
##   <fct>    <fct>    <dbl>    <dbl> <fct>
## 1 Zona Norte 4      232.      52 Apartamento
## 2 Zona Norte 4      272.     160 Casa
## 3 Zona Norte 4      255.     108 Apartamento
## 4 Zona Sur   4      258.      96 Apartamento
## 5 Zona Norte 4      250.      82 Apartamento
## 6 Zona Norte 4      261.     117 Casa
## 7 Zona Norte 4      247.      75 Apartamento
## 8 Zona Norte 4      222.      60 Apartamento
## 9 Zona Norte 4      227.      84 Apartamento
## 10 Zona Norte 4      255.     117 Apartamento
## # i 1,677 more rows
```