

# Full Data - Análisis Exploratorio de Datos

## 1. Análisis de Tendencia Central y Dispersión:

Al analizar las variables en Orange con Feature Statistics, vimos lo siguiente:

- El Ingreso Mensual tiene una media de alrededor de 91.700 y una mediana cercana, lo que muestra que no hay tanta diferencia entre el promedio y el valor central.
- El Gasto Mensual ronda los 62.000 tanto en media como en mediana, lo cual es consistente con lo que esperaríamos de hogares que destinan buena parte de sus ingresos a gastos.
- La Edad Promedio del Hogar está cerca de los 40 años, con una dispersión moderada, lo que refleja que hay hogares jóvenes y otros más grandes.

En general, los valores son coherentes con un contexto económico real: los gastos representan una proporción importante de los ingresos y la edad promedio de los hogares se ubica en la adultez.

## 2. Verificación de Máximos y Mínimos:

En IngresoMensual, los valores van de \$40.607 a \$148.557. Aunque es un rango amplio, puede considerarse realista porque refleja distintos niveles de ingreso entre hogares.

En GastoMensual, el mínimo es \$20.623 y el máximo \$127.691. El gasto más alto es elevado y cercano al ingreso máximo, lo que llama la atención, pero no es imposible.

En EdadPromedioHogar, se detecta un valor mínimo de -5, que claramente no tiene sentido y corresponde a un error de registro. El máximo de 120 años es raro, pero podría darse como caso extremo en algún hogar con personas muy mayores.

En conclusión, salvo la edad negativa, los demás valores son razonables aunque algunos extremos parecen poco comunes.

## 3. Relación entre Variables

Al analizar la relación entre la EdadPromedioHogar y las variables IngresoMensual y GastoMensual, no se observa una tendencia clara o lineal. Los puntos están bastante dispersos, lo que indica que la edad promedio del hogar no determina de forma evidente cuánto ingresan o gastan.

De todas formas, se puede ver que en general los hogares, sin importar la edad promedio, tienden a concentrarse en un rango medio de ingresos y gastos, con algunos valores extremos. Esto sugiere que factores como la zona de vivienda o el tipo de hogar podrían tener más influencia que la edad en sí.

## 4. Análisis de Outliers

En los boxplots vimos que en IngresoMensual y GastoMensual aparecen valores muy altos respecto al resto, que se pueden considerar outliers. También en EdadPromedioHogar se detectan casos atípicos, como edades negativas o mayores a 100 años. Estos valores son inusuales porque se alejan mucho de la distribución normal de los datos, y pueden explicarse por errores de carga o por hogares con situaciones muy particulares.

## 5. Comparación entre Grupos

Al comparar en los boxplots la variable EdadPromedioHogar entre los distintos tipos de vivienda, no se notan grandes diferencias entre casas y departamentos: en ambos casos la edad promedio de los hogares se concentra entre 30 y 50 años. En cambio, cuando lo miramos por zonas de vivienda, sí aparecen algunas variaciones: en la zona rural hay mayor dispersión y algunos valores extremos más altos, mientras que en las zonas urbanas y suburbanas las edades se concentran más cerca del promedio. Esto puede sugerir diferencias demográficas según el contexto de vivienda.

## **6. Correlaciones**

En el heatmap de correlación se observa que la relación más fuerte es entre IngresoMensual y GastoMensual, lo cual tiene sentido porque a mayores ingresos suele haber mayores gastos. También aparece cierta relación positiva entre NumPersonas y GastoMensual, indicando que en hogares más grandes el gasto tiende a aumentar. En cambio, la EdadPromedioHogar no muestra correlaciones fuertes con ingresos ni con gastos, lo que sugiere que la edad de los integrantes del hogar no está directamente vinculada con esas variables.